

社会科学量化研究方法系列

结构方程模型 的原理与应用

邱皓政 林碧芳 著

 中国轻工业出版社

图书在版编目 (CIP) 数据

结构方程模型的原理与应用 / 邱皓政, 林碧芳著. —

北京: 中国轻工业出版社, 2009.2

(社会科学量化研究方法系列)

ISBN 978-7-5019-6725-4

I. 结… II. ①邱…②林… III. 社会科学—统计
模型: 线性模型—研究 IV. C32

中国版本图书馆 CIP 数据核字 (2008) 第 181843 号

总策划: 石 铁

策划编辑: 徐 玥

责任终审: 孟寿萱

责任编辑: 徐 玥

责任监印: 吴维斌

出版发行: 中国轻工业出版社 (北京东长安街 6 号, 邮编: 100740)

印刷: 北京天竺颖华印刷厂

经 销: 各地新华书店

版 次: 2009 年 2 月第 1 版第 1 次印刷

开 本: 787 × 1000 1/16 印张: 29

字 数: 450 千字

书 号: ISBN 978-7-5019-6725-4/C · 021 定价: 52.00 元

咨询电话: 010-65595090 65262933

读者服务部邮购热线电话: 010-65595091 65241695 传真: 85111730

发行电话: 010-65128898 传真: 85113293

网 址: <http://www.chlip.com.cn>

E-mail: club@chlip.com.cn

如发现图书残缺请直接与我社读者服务部 (邮购) 联系调换

81034J6X101ZYW

■推荐序

统计方法的进步,在计算机技术发达的今天尤其神速,新的理论与方法不断被提出。这不但丰富了统计科学本身,更扩展了社会科学等应用研究者的视野。一些基于不同原因而无法跟随当代技术与与时俱进的研究者们,则逐渐落后于创新的步伐,逐渐退出学术的舞台。与其说这是社会进化发展的一种现实,不如说是整个学术界得以蜕变跃进的关键契机。

就我在大学讲台授课以来的一甲子光阴中,确实见证了这个变迁的脉络。从统计学来说,从最简单的次数分布表与百分比,到平均数的 t 检验与实验方法所依据的方差分析,一直到多元变量的因素分析与回归分析,这些方法与技术虽已逐渐成为国内高校教材中的标准课题,但更高级的统计模型原理与技术也应纳入学生学习的范围。另外就测验领域来说,传统以经典测量理论所发展出来的测验编制技术,也逐渐扩展得更精细、更宏观(如项目反应理论、结构方程模型等)。尤其是 20 世纪 70 年代所发展出来的结构方程模型,到今天已经是一套完整的统计模式,能够整合潜变量的界定与测量,分析复杂变量结构,处理多重抽样结构数据,可以有效解决研究者面对不同研究课题的多样性需求,从而被视为一门统计方法学,在过去几十年间独领风骚。关于结构方程模型的原理与应用,邱皓政教授的书中有详细的说明。近年来,结构方程模型结合非线性模型与多层次模型的概念,正向着更一般化的方向发展。

在教育与管理领域进行的研究,除了面对潜变量测量与分析之外,另一个现实是研究者所面对的总体带有层级结构,例如教育系统中的学生-教师-学校-省市-国家,管理研究中的员工-主管-部门-公司-地区-产业,因此研究者在规划研究课题时,就必须具备多层次数据分析的思维与设计,注意样本的独立性假设是否成立。善用这种模型的研究者,将可以从同一批数据中有更多的发现,有兴趣的读者可以在温福星教授的著作中得到这方面的详细说明。

对于一门新兴学问的生根发展,除了研究人员深入持续的基础研究、探索、验证外,还要能够与教育实践相结合,加以推广。因此,教材发展可以说是重要的工作。过去多年来,在邱教授积极热心地不断努力下,台湾统计方法学学会的核心成员多次来访,进行讲学、座谈与合作研究,产生了很大的影响。在课堂中,邱教授和温教授与师生的交流是面对面的,如今,邱教授《结构方程模型的原理与应用》与温教授的《阶层线性模

型的原理与应用》两本书的出版,使得知识的传递与影响更加广泛。

除了研究著说的成就,邱教授与温教授两位最值得称道之处,是能够不断保持学习与创新。每一年他们到北京来时,总能带来一些新东西、新想法或新作品,他们自己戏说是来“交作业”,但重要的是做学问的人总要不断进取,成果才能质量兼备。两位教授身上所体现的做学问的态度,是任何做学问者的必要特质。两位教授今日的成就与影响力,除了两位本身的天赋之外,更是这些特质所造就的。所以,当两位来找我为他们的著作写序时,我欣然同意。但愿我的几句话能够对他们有些许帮助,对中国年轻一代的学术发展有所启发。同时也期望他们乃至其他台湾学者有更多的优秀作品在大陆问世,为华人的学术发展做出贡献。

张厚粲

2008年11月10日

于北京师范大学

■ 自序

关于这本书，让我来说个故事。

1997年，刚回台湾教书的我参加了第二届华人心理学家会议，所报告的论文内容是把结构方程模型（Structural Equation Modeling, SEM）应用在效度的跨样本恒等性比较。犹记得会场中许多海峡两岸的知名教授云集，对于初出茅庐的我，在那种大场合当中可真是名不见经传的人微言轻。走路往往走在最后，吃饭总是躲在角落，但却落得清闲自在。

某天中午，我到餐厅晚了，连个角落的位置都没有，便硬着头皮端着盘子走到一个空位，打断旁边正在谈话的一群教授：“请问我可以坐在这里吗？”即刻我听到一句非常响亮的京片子：“当然可以”，出自一位脸上堆满和蔼笑容的教授口中。稍后不久，她转头问起我打哪里来，做什么的。我简短回答她之后，她居然说我所研究的东西对大陆与台湾都很新颖、很重要，勉励要我好好做，有空到北京来找她。我心里一方面诧异一方面高兴，觉得有人跟我一样重视结构方程模型，让我精神大振，但是惭愧的是，我并不知道她是谁，只约略听到别人称呼她“张先生”。

会后，我遇到一位上海师范大学的学生，问起她我所巧遇的“张先生”是谁，她听完我的描述后，瞪大眼睛说：“该不会是张厚粲教授吧？”我还是摸不着头脑地问她：“张厚粲是谁？”她正色道：“我们都是读她的书长大的，在测验统计领域，她是地位崇高、令人景仰的学者。”她还好奇地问我跟“张教授”是什么关系。我笑了笑，没回答她。

2001年11月，北京师范大学心理学院成立，我随同台湾辅仁大学心理系访问团拜访北师大。在欢迎餐会上，我再次坐在“张先生”身旁。这一回，我已经知道她的身份与地位，本想跟她好好畅谈SEM，没想到除了见识到张先生喝二锅头的“功力”之外，根本没有私下交谈的时间。只记得我跟她说，我写了一本SEM的书，改天给她送去，她很认真地跟我说：“小老弟，辅仁的统计测验就看你的了，大陆的发展重责大任也要你接下去。”然后，又干了一杯二锅头。

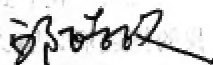
我是地道的台湾人。在1997之前完全与北京没有任何渊源，与大陆没有任何接触。但是今天我会喜欢上北京的二锅头，对大陆的统计测验界的发展非常关心，都是因为张先生。每一次与张先生谈话，都会让我热血沸腾、情绪激昂，并进而转化成具体的行动。或许是因为我以张先生的小小小辅仁师弟自居，更重要的是，我深深地被

IV 结构方程模型的原理与应用

她的风范所吸引、被她的思想所折服。从 2004 年开始,不论我再忙,每年都会来到北京,与北师大学生聊聊、与中国科学院师生交流或到北师大珠海分校、清华大学与北京大学拜访。

科学是逻辑理性明晰的,感情却是隐晦细腻绵密的;这两样,却都是这本书出版的基础。1992 年,我被 SEM 吸引而到美国攻读学位。这套技术果真丰富了我的学术生命,让我能在台湾的学术领域占有一席之地,但让它产生深远意义的更是在大陆这片土地。我永远无法忘记,那年第一次在北师大图书馆的三天 SEM 讲座,台下几百张年轻人认真学习的面庞,以及他们眼中所绽放的光彩。张先生那样的忙碌,却亲自主持我与学生的座谈。让一位来自遥远南方的读书人,不会畏惧北方的风雪冰寒,追求真理真知的力量,源源不绝。

1997 年,在香港中文大学餐厅的一番对话引我踏进一个崭新的世界,坐拥一个开阔的天地。过去多年来张先生对我说过许多话,我未必一一记得;但是我一直知道不要懈怠自己,也不要妄自菲薄。体育界有西方的奥林匹克精神,学术界更有中国自古以来的文人志气,这都要靠我们一代一代坚持传递下去。我尽一己之力,不但丰富了自己,更丰富了知识分子所共有的历史生命。这一本书最终能够在大陆出版,再一次地履行我心中的承诺,也算是那一回餐厅里知遇之恩的回报,更是新旅程的开始。让我们一起面对挑战,也一起享受路上的风光气象。



谨志于台湾

2008 年 8 月

万千心理图书目录

代号	书 名	译 者	定价
心理学导读系列			
X572	心理学导论——思想与行为的认识之路 (第11版)	郑 钢 等译	98.00
X602	心理资本——打造人的竞争优势	李超平 译	32.00
X594	心理学与个人成长	胡佩诚 等译	42.00
X352	人格心理学 (第六版)	陈会昌 等译	48.00
X408	发展心理学 (第六版)	邹 泓 等译	75.00
X300	教育心理学	张厚粲 译	60.00
X620	职业心理学——平衡你的工作与生活 (第七版)	姚 翔 等译	58.00
X567	生理心理学 (第六版)	苏彦捷 等译	60.00
Q568	性与生活——走近人类性科学	甄宏丽 等译	59.00
X338	工业与组织心理学 (第八版)	时 勘 等译	42.00
X531	社会心理学 (第五版·中文第二版)	侯玉波 等译	58.00
X501	异常心理学 (第四版)	杨 霞 等译	80.00
X583	媒介心理学	相德宝 译	42.00
X595	儿童绘画心理学	李 甦 译	50.00
X601	积极心理学	郑 雪 等译校	34.00
X633	调查研究手册	王 彦 译	36.00
X618	质性研究设计	陈 浪 译	14.00
X622	态度与行为——理论、测量与研究	何清华 等译	22.00
X447	心理学批判性思维	李小平 等译	26.00
X448	临床心理学 (第六版)	丛 中 等译	60.00
X454	实用心理评估	张厚粲 编著	24.00
X466	认知心理学 (第三版)	杨炳钧 等译	55.00
X467	情绪心理学 (第五版)	王 力 等译	36.00
X468	心理学研究手册	周晓林 等译	40.00
X469	健康心理学 (第四版)	胡佩诚 等译	65.00
X470	实验心理学教程——勘破心理世界的侦探	郭秀艳 等译	40.00
X339	实验心理学——通过实例入门	张 奇 等译	28.00
X479	学会提问——批判性思维指南	赵玉芳 等译	18.00
X353	问题解决心理学	张 奇 等译	20.00
X409	与“众”不同的心理学	范 照 等译	20.00
X307	改变心理学的40项研究	白学军 等译	26.00

X053	犯罪行为心理学	吴宗宪 等译	50.00
X046	三种心理学——弗洛伊德、斯金纳和 罗杰斯的心理学理论	石 林 等译	13.00
X047	弗兰克尔：意义与人生	常晓玲 等译	18.00
X515	现代英汉—汉英心理学词汇（修订版）	张厚粲 等编	35.00
X058	互联网心理学	谢 影 等译	16.00
X164	现代精神分析“圣经”——客体关系与自体心理学	贾晓明 等译	16.00

心理学导读系列合计

1516.00

心理咨询治疗系列

X332	行为矫正——原理与方法（第三版）	石 林 等译	50.00
X340	心理障碍临床手册（第三版）	刘兴华 等译	78.00
X631	心理诊断和治疗手册——给心理治疗师的指南	卢 宁 等译校	62.00
X634	心理治疗师培训教程	刘 盈 等译	40.00
X453	临床心理督导纲要（第三版）	王择青 等译	56.00
X343	心理咨询师的问诊策略（第五版）	张建新 等译	80.00
X486	抑郁情绪调节手册	汤 臻 等译	38.00
X344	危机干预与创伤治疗方案	梁 军 译	24.00
X345	心理咨询与治疗的理论及实践（第七版）	石 林 等译	40.00
X346	心理咨询与治疗经典案例（第六版）	石 林 等译	30.00
X612	儿童心理测验——更好地理解孩子	傅 莉 等译	28.00
X626	儿童青少年临床访谈技术	徐 洁 译	30.00
X632	儿童游戏——在游戏中成长	谭 晨 译	30.00
X625	脑锁——如何摆脱强迫症	谢际春 等译	30.00
X587	自我催眠	方 新 译	15.00
X588	解读自杀心理	聂 晶 译	32.00
X605	日益亲近	童慧琦 译	36.00
X333	精神分析案例解析	钟 慧 等译	24.00
X347	家庭治疗技术	方晓义 等译	24.00
X412	家庭治疗基础	林丹华 等译	39.00
X348	人格障碍的认知治疗	翟书涛 等译	32.00
X377	团体心理治疗——理论与实践	李 鸣 等译	45.00
X371	认知治疗技术——从业者指南	张黎黎 等译	40.00
X406	精神分析导论	施琪嘉 主译	16.00

更多信息请登录：www.wqedu.com

* 本目录定价如有错误或变动，以实际出书定价为准。

联系地址：北京市朝内大街124号 万千心理（邮编：100010） 咨询电话：010-65595090, 65595091

目 录

第一章 结构方程模型概说.....	1
第一节 结构方程模型的特性.....	3
第二节 结构方程模型的执行.....	10
第三节 SEM 的执行重点.....	14
第四节 结 语.....	24
第二章 结构方程模型的组成.....	27
第一节 结构方程模型的变量.....	28
第二节 结构方程模型的参数.....	31
第三节 模型界定.....	39
第三章 参数估计与识别问题.....	43
第一节 模型识别问题.....	44
第二节 参数估计.....	52
第三节 参数估计策略.....	56
第四节 参数估计的相关议题.....	64
第四章 模型拟合评鉴.....	69
第一节 模型评鉴的基本概念.....	70
第二节 模型评鉴的方法.....	73
第三节 结 语.....	89
第五章 验证性因素分析与模型修饰.....	91
第一节 验证性因素分析原理.....	92
第二节 测量模型的内部拟合检验.....	99
第三节 验证性因素分析的实例.....	108

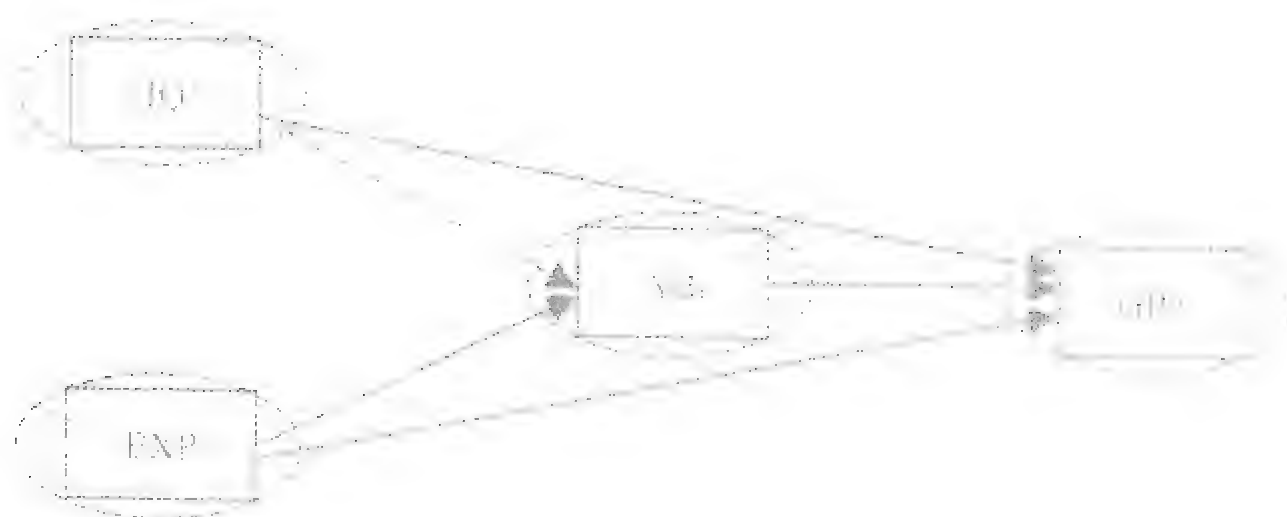
VI 结构方程模型的原理与应用

第四节 验证性因素分析的模型修饰.....	141
第五节 AMOS 的验证性因素分析.....	155
第六节 Mplus 的验证性因素分析.....	165
第七节 结 语.....	170
第六章 高阶验证性因素分析.....	171
第一节 高阶验证性因素分析的概念.....	173
第二节 高阶验证性因素分析的实例.....	176
第三节 AMOS 的高阶验证性因素分析.....	190
第四节 Mplus 的高阶验证性因素分析.....	194
第七章 路径分析.....	197
第一节 路径分析的基本概念.....	198
第二节 路径分析的模型界定与识别.....	204
第三节 LISREL 的路径分析.....	221
第四节 AMOS 的路径分析.....	245
第五节 Mplus 的路径分析.....	251
第六节 结 语.....	255
第八章 结构方程模型：统合模型分析.....	257
第一节 统合模型的基本概念.....	258
第二节 统合模型的分析步骤.....	263
第三节 变量组合与聚合.....	265
第四节 LISREL 的统合模型分析.....	269
第五节 AMOS 的统合模型分析.....	289
第六节 Mplus 的统合模型分析.....	293
第七节 结 语.....	297
第九章 多样本结构方程模型.....	299
第一节 多样本分析的概念.....	300
第二节 多样本分析的统计原理.....	305
第三节 多样本分析：测量恒等性检验.....	311

第四节 AMOS 的多样本分析.....	324
第五节 Mplus 的多样本分析.....	332
第六节 多样本分析: 复核效化检验.....	335
第七节 结 语.....	344
第十章 平均数结构分析.....	345
第一节 平均数结构分析的原理.....	346
第二节 平均数结构的分析技术.....	349
第三节 平均数结构分析: 测量模式.....	351
第四节 平均数结构分析: 统合模型.....	369
第五节 结 语.....	389
第十一章 结构方程模型的正确运用.....	391
第一节 正确运用 SEM 的相关议题.....	392
第二节 SEM 的解释与应用.....	395
第三节 结语: SEM 的展望.....	398
附录一 LISREL 语法.....	401
附录二 SIMPLIS 语法.....	421
附录三 LISREL 常见问题.....	431
附录四 SEM 操作常见疏失检核表.....	435
参考文献.....	437
中文部分.....	437
英文部分.....	437
英汉术语对照表.....	449

第一章

结构方程模型概说



2 结构方程模型的原理与应用

如果问过去的一个世纪中,影响社会科学研究最大的统计方法是什么,想必非因素分析(factor analysis)与回归分析(regression analysis)莫属。1904年,Spearman提出了心理特质的潜在结构因素分析模型,开启了潜在变量模型的大门。另一方面,Wright则自1918年开始,将回归分析扩大到多重联立方程式的估计,正式将回归分析提升到路径模型的层次。一百年来,潜在变量模型与路径分析这两个重量级的量化范式各拥其主,蓬勃发展,但却始终没有交集。直到20世纪70年代,Jöreskog(1973)利用数学矩阵的观念将两种范式巧妙整合,开创了一个崭新的量化研究范式,正式宣告结构方程模型(SEM)的时代来临。

就方法层面来说,结构方程模型是一门基于统计分析技术的研究方法学(statistical methodology),可用来处理复杂的多变量研究数据的探究与分析。重要的是SEM能够同时进行潜在变量的估计与复杂自变量/因变量预测模型的参数估计,也因此被归为多变量统计(multivariate statistics)的一环¹。自从Jöreskog(1973)提出最大似似参数估计的SEM分析概念,并发展出LISREL软件来进行繁复的计算程序之后,SEM已成为当今社会与行为科学计量研究中最重要的一個统计方法范式(Byrne, 1994)。到现在,SEM不仅拥有专属的期刊《结构方程模型》(*Structural Equation Modeling*)、多种专门的统计软件,SEM的研究者也结成一个网络社群SEMNET,一个由70余个国家与地区将近1500位学者专家所结合而成的虚拟化研究社群²。

¹ 美国心理学会于2000年所出版的《Reading and Understanding More Multivariate Statistics》中,有两个专门章节在介绍SEM;此外,由Tabachnick与Fidell所撰写的多变量分析的经典著作《Using Multivariate Statistics》,于第四版(2001)亦以专章详细介绍SEM的原理与应用。

² SEMNET全名为The Structural Equation Modeling Discussion Network,成立于1993年2月,是一个由美国阿拉巴马大学Seebeck计算机中心所支持成立的虚拟网络社群,并以电子邮件传递讯息。有兴趣的读者可以至<http://www.gsu.edu/~mkteer/semnet.html>网页浏览相关讯息或订阅电子报。

第一节 结构方程模型的特性

一、外显与潜在变量

SEM 一个重要的特性是能够对抽象的构念 (construct) 进行估计与检定。社会与行为研究经常必须处理一些抽象的概念, 例如我们想知道“你幸福吗? ”。此时, 针对抽象的概念, 我们必须给予一个操作化的定义, 以便能够透过该程序得到具体的数据, 用以反映不同个体在该概念上的强度。此时, 我们所具体进行测量的变量称为外显变量 (manifest variable)、观察变量 (observed variable) 或测量变量 (measured variable)。这些可以直接获得数据的变量, 若是受到同一个潜在构念的影响, 则会具有共同性, 反映在变量之间的共变关系上。如果针对这些变量之间的共同性加以估计, 得到的一个能够反映该潜在构念的强度的数据, 称之为潜在变量 (latent variable)。

在研究中具体用来获得潜在变量的策略, 最典型的代表是因素分析¹ (factor analysis)。因素分析系利用一组测量同一个构念的观察变量来估计背后的潜在变量。传统因素分析针对于观察变量的背后具有哪几个潜在变量以及潜在变量与观察变量之间的关系为何。它无法事前预知, 直到数据搜集完成之后, 进行变量间的共变关系分析, 抽取出最适当的因素, 确立一个最佳的因素结构模型, 将潜在变量加以命名。此一程序所进行的因素分析, 得到的潜在因素即是一种经验性的潜在变量, 因而被称之为探索性因素分析。

相对之下, 在结构方程模型之中, 潜在变量的概念与内涵是基于理论的推导, 且潜在变量与观察变量的关系是在资料搜集完成之前即事先提出的假设性概念。然后透过实际所搜集的数据, 分析比对假设模型与观察到的数据之间的差异性, 决定研究者对于潜在变量所提出的假设性看法是否恰当。以此种模式进行的因素分析称为验证性因素分析, 也即是一种先验性、事前的潜在变量定义模式。

¹ 因素分析又可译为因子分析, 其中的因素 (factor) 也可译为因子, 表示潜在变量或构念。

二、变量关系的探究

社会及行为科学研究的变量关系，通常并不是单纯的一个变量的推论或两变量关系的讨论，而是涉及一组变量之间关系的讨论。这一组变量除了存在有数学的、表面上的关系外，可能还存有潜在的因果性（causality）或阶层性（hierarchy）。例如一项对于学业表现的研究，最常用的解释变量为智力。然而，研究者可能会考虑到除了智力因素外，学生先前学习经验（Exp）亦会影响学生学业表现（GPA）。而先前学习经验的影响，还可能基于成就动机（Ach）的中介作用间接影响到学业成就。从上述概念中，可以得到如下的研究等式：

$$GPA = a \times IQ + b \times Exp + c \times Ach + e_1 \quad (1.1)$$

$$Ach = d \times Exp + e_2 \quad (1.2)$$

方程式 1.1 与 1.2 陈述了智商、学习经验、成就动机与学业表现四个变量相互影响作用的途径，也就是变量的结构性关系。若以图形表示，方程式 1.1 与 1.2 可绘制成图 1.1 的路径图（path diagram）。

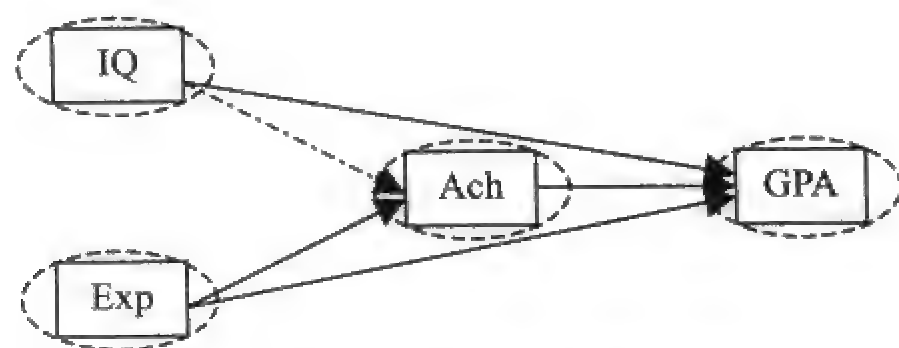


图 1.1 四个变量的假设关系

图 1.1 当中的实线反映了公式 1.1 与 1.2 的关系，但是由 $IQ \rightarrow Ach$ 的关系并未在两个方程式中定义，因此在图 1.1 当中不应该标示出来，而仅以虚线表示。但是如果过去的文献或理论认为智力会影响成就动机，那么 $IQ \rightarrow Ach$ 即必须加以检验，而在图 1.1 当中改以实线表示。换言之，如果研究者调整变量之间的假设关系或扩增研究变量，所需

检验的变量间关系将更为复杂,想要获得支持研究正当性的理论基础与逻辑推理的愿望也就更为殷切。

此外,如果再把潜在变量的观念加以考虑,也就是将图 1.1 当中的各个变量改以椭圆形外框表示,每一个潜在变量由一组观察变量所萃取得出,整个模型即成为一个典型的 SEM 模型。在社会与行为科学领域所探究的变量结构性关系,大多是由一群无法直接观察与测量的抽象命题或构念所组成。此时必须先针对这些构念的定义运用验证性因素分析来确认每一个构念内在结构的適切性,然后进一步检验潜在变量间的结构关系。此点也是 SEM 的主要长处之一 (Bollen, 1989)。

总之,不论是变量因果关系的证明或构念内在结构的确认,均有赖于事前研究变量的性质与内容的厘清,并清楚描述变量的假设性关系,藉由研究者提出具体的结构性关系的假设命题,寻求统计上的检证来确认。整个过程即是结构方程模型。

三、模型比较分析

SEM 的另一个特征是模组化分析的应用。利用先前所讨论的假设检验与结构化验证功能,结构方程模型可以将一系列的研究假设同时结构成一个有意义的假设模型 (hypothetical model),然后经由统计的程序对于此一模型进行检证。不同的模型之间,则可进行竞争比较。

在社会与行为科学的研究中,往往相同的一组变量会因为理论观点的不同,对于变量之间的假设关系亦会有不同的主张。因此,研究者可以基于不同的理论与假设前提,发展出不同的替代模型 (alternative model),进行模式间的竞争比较。此一利用假设模型进行统计检证的优点,大大改善了传统路径分析在多组回归等式进行同时估计的限制,也提高了分析的应用广度。

Jöreskog 与 Sörbom (1996) 指出 SEM 的模组化应用策略有三个层次:第一是单纯的验证 (confirmatory),也就是针对单一的先验假设模型,评估其適切性,称为验证型研究;第二是模型的产生 (model generation),其程序是先设定一个起始模型,再与实际观察数据进行比较之后,进行必要的修正,反复进行估计的程序以得到最佳拟合的模型,称为产生型研究;第三是替代模型的竞争比较,以决定何者最能反

映真实资料，称为竞争型研究。

Maccallum 与 Austin (2000) 从文献整理中发现，以单纯的验证与模型产生为目的 SEM 研究分别约占 20% 与 25%，涉及竞争比较的 SEM 研究则有 55%。Maccallum 与 Austin (2000) 认为模型产生型 SEM 研究有其限制存在，尤其在模型修饰的过程中，往往过度依赖数据所呈现的讯息而忽略理论的意义。过度滥用修正程序以获得对自己有利的结果，是相当危险的做法，使用者应小心为之。相对之下，竞争比较的研究则有较为强固的理论基础，修饰问题较少，而可以发挥较大的弹性与说服力。

结构方程模型的此一模组化分析功能，最主要的一个贡献即是为社会与行为科学研究界对于抽象理论进行实证的检验提供了一套严谨的程序，使得研究者可以透过统计的分析去检验所提出的理论模型 (theoretical model)。此举将假设检验的运用，自单一参数的考验提升至理论模型整体考验的更高层次，突破了传统上计量技术对于理论模型欠缺整合分析能力的困境。

四、结构方程模型的技术特性

Hoyle (1995) 指出，结构方程模型可视为不同统计技术与研究方法的综合体。从技术的层面来看，SEM 并非单指某一种特定的统计方法，而是一套用以分析共变结构的技术的整合。SEM 有时以共变结构分析 (covariance structure analysis)、共变结构模式 (covariance structure modeling) 等不同的名词存在，有时则单指因素分析模式的分析，以验证性因素分析 (CFA) 来称之；有时，研究者虽然以 SEM 的分析软件来执行传统的路径分析，进行因果模型 (causal modeling) 的探究，但不使用 SEM 的名义。事实上，这也是 SEM 的重要应用之一。不论是用何种名词称呼，这些分析技术都具有一些基本的共同特质 (Kline, 1996)，说明如下。

(一) SEM 具有理论先验性

SEM 分析最重要的一个特性，是它必须建立在一定的理论基础之上。也就是说，SEM 是一个用以检证某一先期提出的理论模型 (priori theoretical model) 的適切性的一

种统计技术。这也是 SEM 被视为是一种验证性 (confirmatory) 而非探索性 (exploratory) 统计方法的主要原因。SEM 的分析过程中, 从变量内容的界定、变量关系的假设、参数的设定、模型的安排与修正, 一直到应用分析软件进行估计, 其间的每一个步骤都必须要有清楚的理论概念或逻辑推理作为依据。从统计的原理来看, SEM 也必须同时符合多项传统统计分析的基本假设 (例如线性关系、正态性) 以及 SEM 分析软件所特有的假设要件, 否则所获得的统计量据无法采信。

以因素分析为例, 结构方程模型所使用的因素模式采取了相当严格的限制。研究者对于潜在变量的内容与性质, 在测量之初即必须有非常明确的说明, 或有具体的理论基础, 并已先期决定相对应的观察变量的组成模式。分析的进行即在考验这一先期提出的因素结构的适切性。除了测量工具发展时, 可以利用此一程序检验其结构的有效性, 也可用于理论架构的检验, 因此又被称为验证性因素分析。

(二) SEM 同时处理测量与分析问题

过去传统的统计方法, 不论分析的内容为何, 多把变量视为“真实”、“具体”、“可观测”的测量数据。在分析过程中, 并不去处理测量过程所存在的问题, 也就是说, “测量”与“统计”是两个独立分离的程序。传统上, 如果变量所涉及的概念是如同“智力”或“焦虑”等不易界定的心理概念, 研究者为了获得可以分析的数据, 会先行讨论测量的方法, 并以信度与效度的概念程序先行进行评估。一旦通过评估的标准, 即将所获得的测量数据进行分析。

相对于传统的做法, SEM 是一套可以将“测量”与“分析”整合为一的计量研究技术。主要的关键在于 SEM 将不可直接观察的构念或概念, 以潜在变量的形式, 利用观察变量的模型化分析加以估计。不仅可以估计测量过程当中的误差, 也可以用以评估测量的信度与效度 (如因素效度), 甚至可以超越古典测量理论的一些基本假设, 针对特定的测量现象 (例如误差的相关性) 加以检测。另一方面, 在探讨变量之间关系的时候, 测量过程所产生的误差并没有被排除在外, 而是同时包含在分析的过程当中, 使得测量信度的概念可以整合到路径分析等统计推论的决策过程中。

(三) SEM 以协方差的运用为核心，亦可处理平均数估计

SEM分析的核心概念是变量的协方差¹ (covariance)。协方差是描述统计中的一种离散量数，利用方差 (variance) 的离均差和的数学原理，计算出两个连续变量配对分数 (paired scores) 的变异量，用以反映两个变量的共同变异或相互关联程度。协方差是一个非标准化的统计量数，受到两个变量所使用的量尺或单位的影响，数值可能介于 $-\infty$ 到 $+\infty$ 之间，如果将协方差除以两个变量的标准差，即可得出标准化协方差 (即 Pearson 相关系数)。

在 SEM 当中，协方差具有两种功能：第一是描述性的功能，利用变量之间的协方差矩阵，我们可以观察出多个连续变量之间的关联情形；第二是验证性的功能，用以反映出理论模型所导出的协方差与实际观测得到的协方差的差异。分析过程中最重要的数学程序，即是计算导出协方差矩阵 (Σ matrix)。如果研究者所设定的 SEM 模型有问题，或是数据估计过程导致 Σ 矩阵无法导出，整个 SEM 即无法完成。

除了协方差以外，SEM 也可以处理变量的集中趋势的分析与比较，也就是平均数的检验。传统上，平均数检验是以 t 检验或方差分析 (ANOVA) 来进行。由于 SEM 可以对于截距进行估计，使得 SEM 可以将平均数差异的比较纳入分析模型当中，同时若配合潜在变量的概念，SEM 更可以估计潜在变量平均数，使应用范围更为广泛。

(四) SEM 适用于大样本的分析

由于 SEM 所处理的变量数目较多，变量之间的关系较为复杂，因此为了维持统计假设不致违反，必须使用较大的样本数。同时样本规模的大小，也牵动着 SEM 分析的稳定性与各种指数的适用性，因此，样本数的影响在 SEM 当中是一个重要的议题。

与其他的统计技术一样，SEM 分析所使用的样本规模当然是越大越好，但是究竟有没有一个最适规模，则会随着 SEM 模型的复杂度与分析的目的及种类而有相当大的变化。但是，一般来说，当样本数低于 100 之时，几乎所有的 SEM 分析都是不稳定的。Breckler (1990) 曾针对人格与社会心理学领域的 72 个 SEM 实证研究进行

¹ covariance 反映两个变量共同变化的程度，除了译为协方差又可译为共变数。variance 反映某个变量数据离散的变异程度，除了译为方差又可译为变异数，在本书中这两组词汇被视为同义词而交替使用。

分析,样本规模介于40至8650之间,中数为198。有1/4的研究小于样本数500,约20%的研究样本规模小于100。因此,一般而言,大于200以上的样本,才可以称得上是一个中型的样本。若要追求稳定的SEM分析结果,低于200的样本数是不鼓励的。

(五) SEM 包含了许多不同的统计技术

综观统计分析技术的内容,可以概略分为平均数检验的方差分析与探讨线性关系的回归分析两大范畴。事实上,这两者并无本质上的差异,前者可以被归为一般线性模型(general linear model)分析技术,后者则是以变量间的线性关系为分析的内容。随着计算机科技的发展,分析软件功能的提升,使得两种统计模式可以互通,合而为一。

一般线性模型的优点是可以数学方式来整合不同形态的变异来源,可以不断扩充研究者所欲探讨的变量的数目与影响方式,因此一般线性模型逐渐发展出多种多变量统计的概念,例如多变量方差分析(multivariate analysis of variance)。而回归分析在处理变量的弹性与复杂度的优势似乎有凌驾方差分析之势,但是方差分析由于简单清楚的数学原理与容易解释分析的特性,也一直受到研究者的青睐¹。在SEM当中,虽然是以变量的共变关系为主要内容,但由于SEM模型往往牵涉到大量变量的分析,因此常借用一般线性模式分析技术来整合变量,故SEM分析可以说是多种不同统计分析程序的集合体。

(六) SEM 重视多重统计指标的运用

虽然SEM涵盖了多种不同统计技术于一身,但是对于统计显著性的依赖性却远不及一般统计分析,主要理由有三:第一,SEM所处理的是整体模型的比较,因此所参考的指标不是以单一的参数为主要考虑,而是整合性的系数。此时,个别检验是否具有特定的统计显著性就不会成为SEM分析的重点所在。第二,SEM发展出多种不同的统计评估指标,使得使用者可以从不同的角度进行分析,避免过度倚赖

¹有关方差分析与相关/回归分析的优劣与取代性,在心理计量领域曾经有过精彩的对话,方差分析着重于实验研究的应用,重视团体差异,相关/回归分析着重于个别差异的分析,受到非实验研究者的欢迎。这些讨论可以参考Cohen & Cohen (1983)、Licht (1995)、Pedhazur (1997)与Tatsoka (1975)的文章。

单一指标。第三，由于 SEM 涉及大样本的分析，当样本越大，SEM 分析的核心概念卡方统计量的显著性，即受到相当的扭曲，因此 SEM 的评估指数都特意避免碰触到卡方检验的显著性检验。也因为这个原因，在 SEM 分析当中，较少讨论到与统计显著性决策有关的第一与第二类型错误议题，显示了 SEM 技术的优势是在于整体层次（macro-level）而非个别或微观的层次（micro-level）。

第二节 结构方程模型的执行

SEM 分析的基本程序可以概分为模型发展与估计评鉴两个阶段。前者在发展 SEM 分析的原理基础上并使 SEM 模型符合特定的技术要求，此时研究者的主要工作在概念推导与 SEM 分析的技术原理的考虑；后者则通过产生 SEM 的计量数据来评估 SEM 模型的优劣好坏。有关 SEM 分析的执行流程的概念，请参考图 1.2。

一、模型发展阶段

模型发展阶段的主要目的在建立一个适用于 SEM 分析概念与技术需要的假设模型（hypothetic model），牵涉到理论发展、模式界定与模型识别等三个概念。在图 1.2 当中，这三个概念虽然是以连续的流程图来表示，但是三者间的关系只是说明概念发生的先后顺序。在实际操作上，这三个概念的运作则是相互作用的不断往复过程。

首先，SEM 模型的建立必定以理论为基础。所谓的以理论为基础，并不是说 SEM 模型一定必须建立在某一个特定的理论之上，而是强调 SEM 模型的建立必须经过观念的厘清、文献整理与推导或是研究假设的发展等理论性的辩证与演绎过程，最终提出一套有待验证的假设模型。在前面的章节中，已经指出了 SEM 的一个重要特性是理论的先验性，因此，SEM 分析的第一个阶段，主要目的便是在建构 SEM 的理论基础。另外两个概念，模式界定与模型识别，也即是根基于理论性推演过程，将 SEM 模型的理论假设转换成为适当的技术语言。

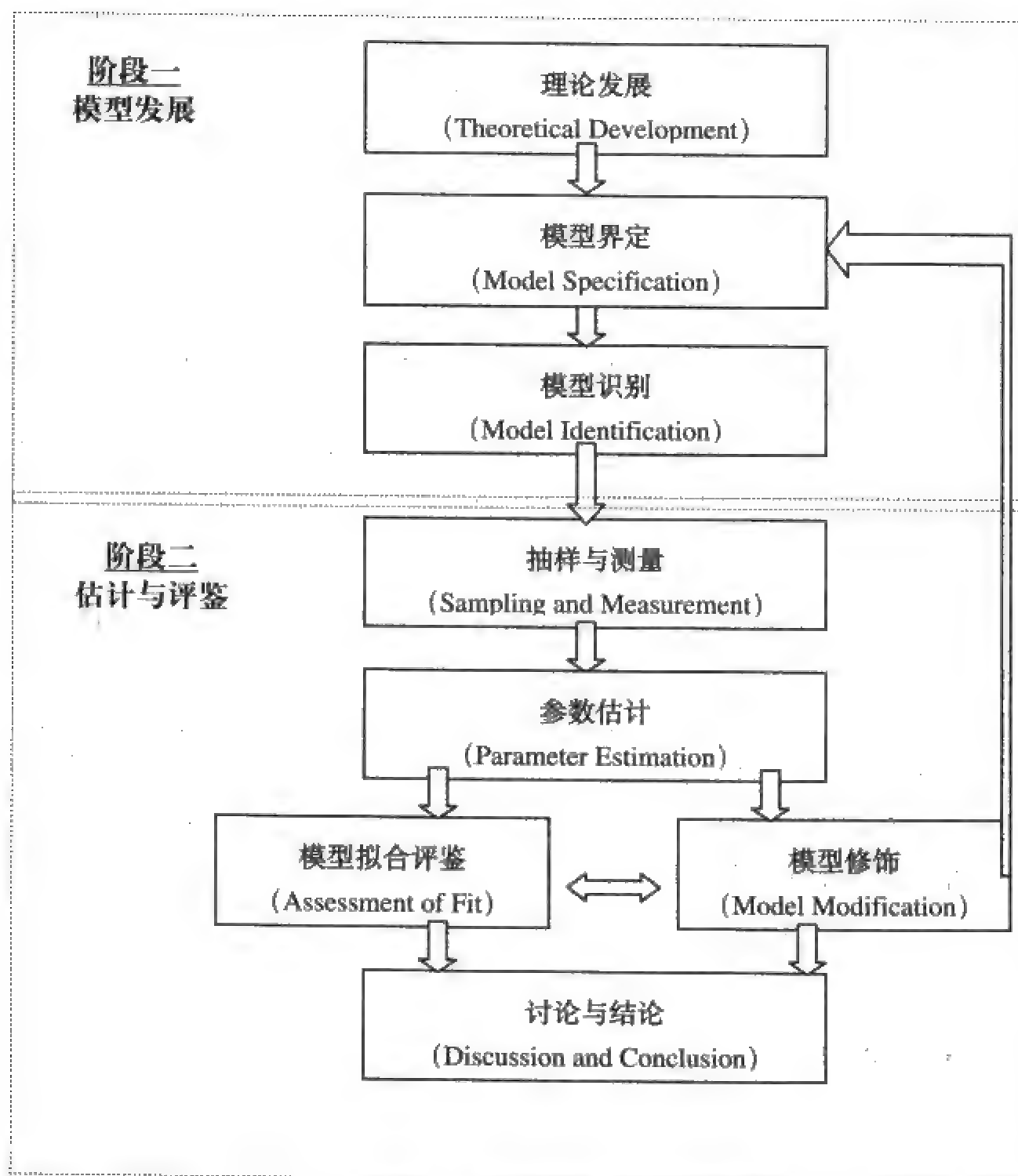


图 1.2 结构方程模型的基本程序

配合理论推导的过程，研究者同时必须进行模型界定的重要工作。模型界定可以说是第一个阶段当中最为具体的步骤，目的在发展可供 SEM 进行检验与估计的变量关系与假设模型。

一般来说, SEM 模型的构成有两个主要的部分:第一是理论或概念的基础,或是研究者个人的先备知识与经验;第二是 SEM 的技术语言与方法要求。当研究者面对所关心的研究问题之时,除了基于自己的知识基础与研究兴趣来推演出值得探讨的研究命题之外,还需对于理论文献详加检阅,以建立严谨的科学假设或提出有待验证的理论模型。

如果研究者选择使用 SEM 来探讨所提出的假设与理论模型,则必须配合特定的 SEM 技术语言与各项操作要求,将研究者所提出的假设与理论模型转换成 SEM 模式。在此同时也需考虑 SEM 分析当中可能涉及的各种统计概念所使用的统计原理,纳入 SEM 模型的设定之中。这个由理论发展到技术性模型建立的一整套程序,就是第二个具体步骤“模式界定”的主要任务。而模式界定的具体产品,是建立一个 SEM 路径图。该路径图就是模型识别步骤据以评估的依据,也是第二阶段进行估计分析的地图。

此外,在模式界定的过程当中,有一个非常重要的技术问题,即必须让 SEM 模型具有可识别性,使 SEM 的各项数学估计程序可以顺利进行。由于结构方程模型所设定的假设模式是基于研究者的研究需求所提出,而模式的分析必须利用实际搜集得到的数据,利用分析软件来进行估计工作;因此,只有在模型符合统计分析 with 软件执行的要求,也就是在能够被有效识别的情况下,SEM 分析才能顺利进行。此时,一个模型可以被有效进行识别的程度称为模型识别度,有关模型识别度的估算过程,是第三个步骤“模型识别”的主要任务。相关议题将在后续章节详细介绍。

二、估计与评鉴阶段

一旦 SEM 模型发展完成之后,研究者即必须搜集实际的测量资料来检验所提出的概念模型的适当性。此一阶段开始于样本的建立与测量工作的进行,所获得观察资料经过处理后,即依照 SEM 分析工具的要求,进行各项估计。

样本的获得对于 SEM 分析的结果有着重要的影响。除了样本规模大小的影响,由于 SEM 涉及潜在变量的测量,因此 SEM 分析的结果与样本结构及测量质量有密切的关系,也就是具有样本的依赖性 (sample realitication)。Bollen (2002) 指出,

对于某些个体而言,潜在变量可能是有意义的概念;但是对另一些个体则可能不然,或有不同的意义;因此在不同的样本间,SEM分析可能得到非常不同的结果。

为了顺应 SEM 会随着样本的特性而改变结果的性质,SEM 的分析工具多提供了评鉴的指标,以反映样本规模与性质的影响。同时 SEM 分析的本身,亦可以处理测量误差的估计,使测量质量的影响可以被有效的控制。但是,研究者仍然必须谨慎选取研究样本,维护测量的质量,因为 SEM 分析的复杂性高,任何一个环节的瑕疵或失误,都可能造成 SEM 分析结果产生变化。

SEM 的参数估计可以说完全是由计算机来进行,只有少数的部分必须由人工计算完成(例如测量模式的信度估计),但是如何让 SEM 分析顺利完成,仍有赖研究者对于计算机软件正确无误的指令下达,以及分析工具的各种选项的正确决定。不同的分析软件各有其优劣,因此操作方法各有不同。SEM 的使用者必须对于 SEM 分析工具的一般性基本原理有深入的了解,也必须从操作演练中累积经验,熟知每一套软件的优劣与限制,了解每一个参数或警告讯息的意义与作用,才能顺利完成各项估计与评估程序。

值得注意的是,在估计与评鉴过程中,SEM 分析工具通常会提供模型调整与修饰的计量信息。使用者可以根据这些指数或统计检验数据,调整先前所提出的假设模型,重新、反复进行估计与模型评估,此一过程称之为模型修饰(model modification)。虽然此一做法违反了 SEM 分析理论先验性的精神,但是观察数据背后所潜藏的各种信息,也是科学研究相当珍贵的线索。从中既可能看出研究者在理论推导过程当中的疏忽或盲点,也可能引导研究者继续推导出更有意义的概念或假设,重新提出一套更趋合理的 SEM 模型。因此,模型的修正步骤也是一般 SEM 使用者相当重视的部分。有关模型修饰的各项指数与方法,我们将在第五章介绍。

第三节 SEM 的执行重点

一、模型的描述与设定

典型的 SEM 分析是由一系列代表特定的研究假设的回归方程式所组成。一个 SEM 模型不仅牵涉到不同类型变量的处理,例如观察变量与潜在变量、内生变量与外源变量等,也涉及不同关系形态的设定,例如回归、相关与残差估计等,因此,SEM 分析首重模型的正确设定与呈现,模型的设定除了依循前面章节的介绍,SEM 分析尤其重视概念路径图的运用(如图 1.2),路径图不仅可以协助研究者呈现他们的研究架构,更可以促成研究者与其他读者之间的沟通。

在 SEM 的应用中,模型有两种呈现形式:概念模型(conceptual model)与统计模型(statistical model)。概念模型主要在说明一个 SEM 研究当中所探讨的概念间的关系,可以利用如图 1.3 的路径图的形式来呈现。该图是由 LISREL 软件的 Path Diagram 指令所产生(点选对话框当中的 conceptual diagram 选项即可获得)。读者可以利用 LISREL 的图形编辑对话框,增加或移除图 1.3 当中模型内的参数路径或变量形态,或是移动位置整理出美观的图形,使用上非常便利。

当研究者在发展概念模型时,并不是要将所有的变量、参数、关系,都无所遗漏地标示在图中,以呈现出一个“看起来”完整的统计模型。当一个模型当中有越多需要估计的参数,该模型越趋于复杂。从简效原则来看,越复杂的模型越不建议采用。Hoyle 与 Panter (1995) 建议概念模型应能适当地反映出研究者所感兴趣的参数或关系,避免漫无目的、缺乏理论根据的绘制模型。具体来说,必须在特定的理论概念基础上,透过研究者的研究假设,才能形成概念模型。

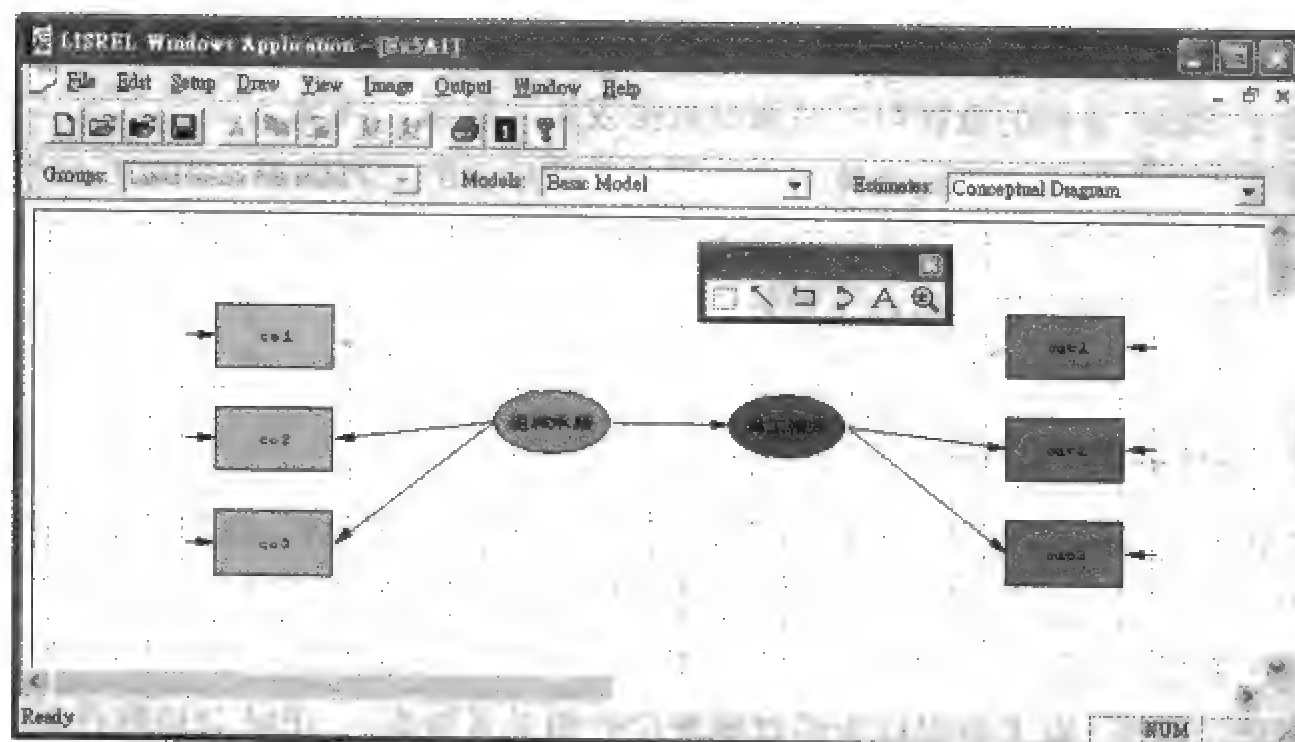


图 1.3 以 LISREL 软件所产生的概念模型

至于统计模型，则是指可以利用分析工具加以检测的操作性模型（operational model）。这里所谓的操作性模型，其意义就好比操作性定义（operational definition）一样，表示如何经过操作的程序以得到概念结果的整个程序。而此时的操作性，除了是指符合统计运算的原理之外，同时也需符合统计分析工具的技术原理与限制。

统计模型与概念模型的差别是概念模型可以利用路径图来具体说明，但是统计模型通常是进行 SEM 分析过程中的一些思考概念与计算原则，而不是具体的一个实体模型。例如，我们要计算模型是否可以顺利地产生参数估计，所以要计算模型的识别度。此时，根据识别原理所发展出来的参数估计安排，就是统计模型的概念，或者，为了要使潜在变量能够量尺化，我们可能会对某一些参数进行特殊的设定，此时也是基于统计模型的考虑。

通常，统计模型最具体的呈现就是语法指令的内容，例如在一个 LISREL 语法当中，就会详细列举哪些参数是被估计参数、哪些参数是固定参数等讯息。然而统计模型的真正内涵，是概念模型得以进行估计分析的整个统计过程。

二、资料的准备

用于 SEM 分析的数据可以是原始数据或矩阵数据。SEM 分析的始祖 LISREL 最早所使用的数据形态即是矩阵。由于影响数据质量的原因很多,因此在进行 SEM 分析之前,也必须先行处理资料完整性与正确性的筛检,使得 SEM 分析可以在一个干净、完整、正确的资料基础上进行。

(一) 矩阵数据的应用

Cudeck (1989) 建议, SEM 分析最好使用的是协方差矩阵,而非相关系数矩阵。一般人或许会以为相关系数可以提供较为清楚的变量关系的描述,因为相关系数是标准化的系数,数据介于+1 与-1 之间,越接近 0 表示关系微弱,越接近 ± 1 表示线性关系明显。甚至,有人会误以为经过标准化的相关系数输入 SEM 分析软件后,会有利于标准化 SEM 参数的标准化数值的估计。

统计学的知识告诉我们,相关系数是将协方差除以标准差所获得。一组变量的协方差矩阵中,不仅可以计算出协方差、方差,也可以计算出相关系数。但是,利用相关系数矩阵进行 SEM 分析,并不能导出协方差的数据,除非我们另行告知 LISREL、Mplus、AMOS 等分析软件有关各变量的标准差数据。也就是说,协方差矩阵能够涵盖相关系数矩阵,最重要的是它能够推导出 SEM 分析所需的各种重要数据,例如变量的方差与协方差等。

在输入矩阵数据时,我们往往会将数据化简到小数点后第二位,但是 Hoyle 与 Panter (1995) 建议应增加至小数点后第三位,以提高计算的精确度。尤其当我们输入相关矩阵与标准差数据时,如果小数点后的位数太少,推导出的协方差估计量的波动性增加,间接造成模型拟合指数的降低,不利于研究结果。一般而言,利用相关矩阵作为输入数据时,在论文中应报告相关矩阵与标准差数据;利用协方差矩阵作为输入数据时,在论文中也应附上矩阵数据,以利查考 (McDonald & Ho, 2002)。

(二) 原始数据的使用

除了矩阵资料之外, SEM 分析也可以直接读取原始资料进行分析。当研究者所

使用的变量包括了类别或顺序变量时,即无法产生协方差矩阵或相关矩阵。此时必须先行处理这些非等距数据,使其在格式上符合 SEM 分析的需求。这时候,原始数据就显得格外重要。

另外,原始数据的使用有助于多阶段 SEM 分析的进行。在许多研究中,一个变量的分数可能是由几个题目的加总测得。在进行 SEM 分析之前,研究者会先行利用其他软件进行加总的工作,再以原始数据或矩阵数据的形态输入计算机中。但是,此一动作也可以透过将加总前的原始数据库,输入 SEM 分析软件后,以多阶层测量模式的估计程序获得因素分数,而不是以加总的方式获得变量分数。在有些 SEM 的分析上,这些以项目层次(item-level)原始数据作为 SEM 分析数据,有其不同的操作程序(例如 MTMM 研究,若以项目层次进行分析,与以总分层次的原始数据输入 SEM 分析,得到的结果可能不同),但是都必须以原始数据作为输入数据。

(三) 变量的分布特性

SEM 分析的进行,必须建立在一定的统计假设基础之上。当统计假设违反时,SEM 分析的数据是值得怀疑的。因此,一般研究论文撰写时,应列举变量的分布特征与统计假设的检验结果(Hoyle & Panter, 1995; McDonald & Ho, 2002)。尤其在当研究者是以矩阵数据作为输入数据时,由于缺乏各变量的原始数据可资判断变量的分布特性,因此,更是需要揭示各变量的次数分布的特征,证明变量的偏态与峰度处在合宜的水平,或是没有遗漏与偏离的状况。

Hoyle 与 Panter (1995) 建议,在撰写研究报告时,应说明变量的正态性、多变量正态性以及峰度的数据。因为某些估计程序明显受到正态性不足的影响,例如 ML 法与 GLS 程序,所以完整清楚地列举检验结果是对研究报告正确性的一个负责任的做法。

三、报表的整理与分析

SEM 分析涉及一连串的参数估计、模型检验与模型修饰的技术处理程序。因此,使用者必须非常熟悉每一个步骤的原理与目的,才能理解报表的内容与分析说明的

重点。

在阅读分析软件的报表时，应分别就两个层次的数据进行处理。第一是过程性的资料，也就是在完成最终结论之前，我们必须详细检阅 SEM 分析的各项数据，观察这些数据的状态，必要时加以记载，作为报告撰写之用。第二是最终解（final solution）的报告，也就是 SEM 分析的最后结论的各种参数数据，以及模型拟合度的最终数据。这两项处理的完成，有下列一些重要的处理原则。

（一）估计方法的选择

在第四章的论述中我们可以知道，SEM 分析可以用不同的估计方法进行参数估计，而不同的方法所得到的结果也有所不同。因此，在 SEM 的研究报告中，应说明使用何种估计策略，并说明为配合该种策略，有无特殊的处理（例如样本规模的决定、变量经过正态校正等），使得读者可以清楚了解 SEM 的各项参数是在何种基础上估计出来的。

一般而言，最常用的估计方法是最大概似法（ML 法）。ML 法的优点是在小样本时，或是变量有不太理想的峰度时，仍然可以获得理想的参数估计数。因此，对于 SEM 分析不熟悉的使用者，可以直接使用 ML 法进行 SEM 分析。

（二）模型拟合指数

模型拟合指数的功能是用来评估一个 SEM 模型是否与观测数据相拟合。在 SEM 的具体应用上，拟合有两种不同的意义：第一是绝对拟合（absolute fit），第二是增量拟合（incremental fit）。前者反映的是模型导出的协方差矩阵与实际观测的协方差矩阵之间的拟合情形。拟合度的数值大小表示模型导出数与实际观测数差异的多寡。至于增量拟合，则是指某一个模型的拟合度较另一个替代模型的拟合度，增加或减少了多少拟合度。例如一个模型假设潜在变量之间具有相关，替代模型则假设潜在变量之间没有相关（称为虚无或独立模型），计算出两个模型的拟合度差异量后，以推知何者较能拟合观测资料。

这两种拟合度的概念适用不同的模型拟合指数。然而所有的 SEM 分析都应先报告卡方统计量以及与卡方统计量的计算有关的讯息（自由度、样本数、显著性数据），

如果是非正态性的数据经过了校正,在报告传统的卡方值之外,也应报告调整后的卡方值 (Scaled χ^2 ; Satorra & Bentler, 1994)。

除此之外, GFI指数可以说是每一个SEM研究都会报告的数据。因为GFI指数性质类似于回归分析的 R^2 ,当数值越大,表示实际观察的协方差矩阵能够被假设模型解释的百分比越高,模型拟合度越佳。GFI指数可以说是反映绝对拟合最佳的指数。

增量拟合的评估可以利用 NNFI、IFI 等指数。这些指数的基础是模型间的卡方值差异值,也就是 Hu 与 Bentler (1995) 所称的第二类型指数。如果研究者使用的是 ML 估计程序, NNFI 或称为 TLI 指数 (Tucker & Lewis's index) 是较常用的指数,但是当样本数少时 (例如低于 150) 则不建议使用,可以改用 IFI 指数。如果研究者采用的是 GLS 估计方法,则以 IFI 指数的表现较理想。

若以非中央卡方为基础比较模型增量拟合,即 Ho 与 Bentler (1995) 所称的第三类型指数,较佳的选择是 CFI 指数 (又称为 BFI 指数, Bentler, 1989; 或 RNI 指数, McDonald & Marsh, 1990); RMSEA 指数则是近年来逐渐受到普遍采用的指数,因为 RMSEA 是第三类的非中央卡方指数当中不受样本分布影响的指数。Hu 与 Bentler (1999) 主张 CFI 与 RMSEA 两个指数都需报告在论文中。当研究者想去估计统计检验力时, RMSEA 是非常适合的指数。当研究者想要比较不同的模型,但是没有嵌套关系时,则可使用 ECVI、AIC 或 CAIC 指数。

在呈现这些数据时,如果分析的模型很多,可以利用表格来整理呈现,以求做到一目了然 (读者可以参考本书范例中的整理方式,或是其他论文的整理格式)。在论文的文字叙述中,可以写为 $\chi^2 (128, N=284) = 506.23, p < .001, NNFI = .89, CFI = .91$ 的形态,在 χ^2 数值呈现时,应一并报告自由度与样本数数据。然后再就数值的内容与意义加以说明。

(三) 参数的报告

当模型拟合指数显示某一个模型是适合的模型之后,研究者应着手整理从该模型估计得出的最终解当中各参数的数据。Hoyle & Panter (1995) 指出参数的报告应该尽可能充分翔实,使得读者可以清楚看出每一个参数的特性与代表的意义,包括

三种重要的讯息：参数的合理性、显著性检验、标准化解，都应完整揭示。

首先，参数估计的合理性反映的是该参数是否符合数学或统计学原理上的可能性，或是实证资料的可能性。一般而言，参数的方差是衡量参数估计数的最重要资料。当残差的方差出现了负值（称为 Heywood 现象），或是出现超过范围的协方差（当标准化的估计数大于 1 时），都表示参数估计是有问题的。在报告 SEM 分析结果时，如果有方差的数值，都应在报告中予以揭示。

其次，是各参数的显著性检验数据应翔实列举。除了指出检验值的大小与显著性以外（例如 t 检验值与 p 值），标准误也是重要的数据，从标准误的大小，可以看出参数估计是否存在潜在的问题。值得注意的是，在 SEM 模型当中，可能有某些参数被设定为固定值（例如被用来作为潜在变量量尺化的因素载荷，通常被设定为 1.00），因此没有经过估计与显著性检验，在说明显著性结果时，亦应予以标注说明。

第三，标准化解的呈现，通常是 SEM 分析最重要的一个部分，因为标准化解反映了 SEM 模型中各参数估计的最后结果，而且是以标准化的形式出现。

一般而言，LISREL 或 EQS 等软件，都会在报表的最后提供标准化解的详细信息，指出相关系数、因素载荷、回归系数等各项数值的结果。以 LISREL 为例，在分析完成之后，可以要求列出标准化解的路径图，以图标法列出所有参数的标准化解，如图 1.4 所示。

（四）标准化解与完全标准化解

值得注意的是，LISREL 分析会产生两种不同的标准化解：以 SS（standard solution）指令所获得的标准化解以及以 SC（solution completely standardized）指令得到的完全标准化解。标准化解的数学原理是针对潜在变量的方差估计数进行标准化，也就是将潜在变量的估计数除以潜在变量的估计标准差（以去除每一个潜在变量各有不同的变异情形的影响），然后计算出所有参数的数值。完全标准化解则是除了将潜在变量的数据加以标准化之外，还将观察变量的估计数加以标准化。也就是将每一个观察变量的数据，除以各变量的标准差（以去除每一个观察变量各有不同的变异情形的影响），然后计算出所有参数的标准化估计数。对于不具有结构模式的

SEM 分析（例如 CFA 分析），SS 与 SC 所得到的数据会完全相等。因为对因素的标准化，也即是对测量变量的标准化。

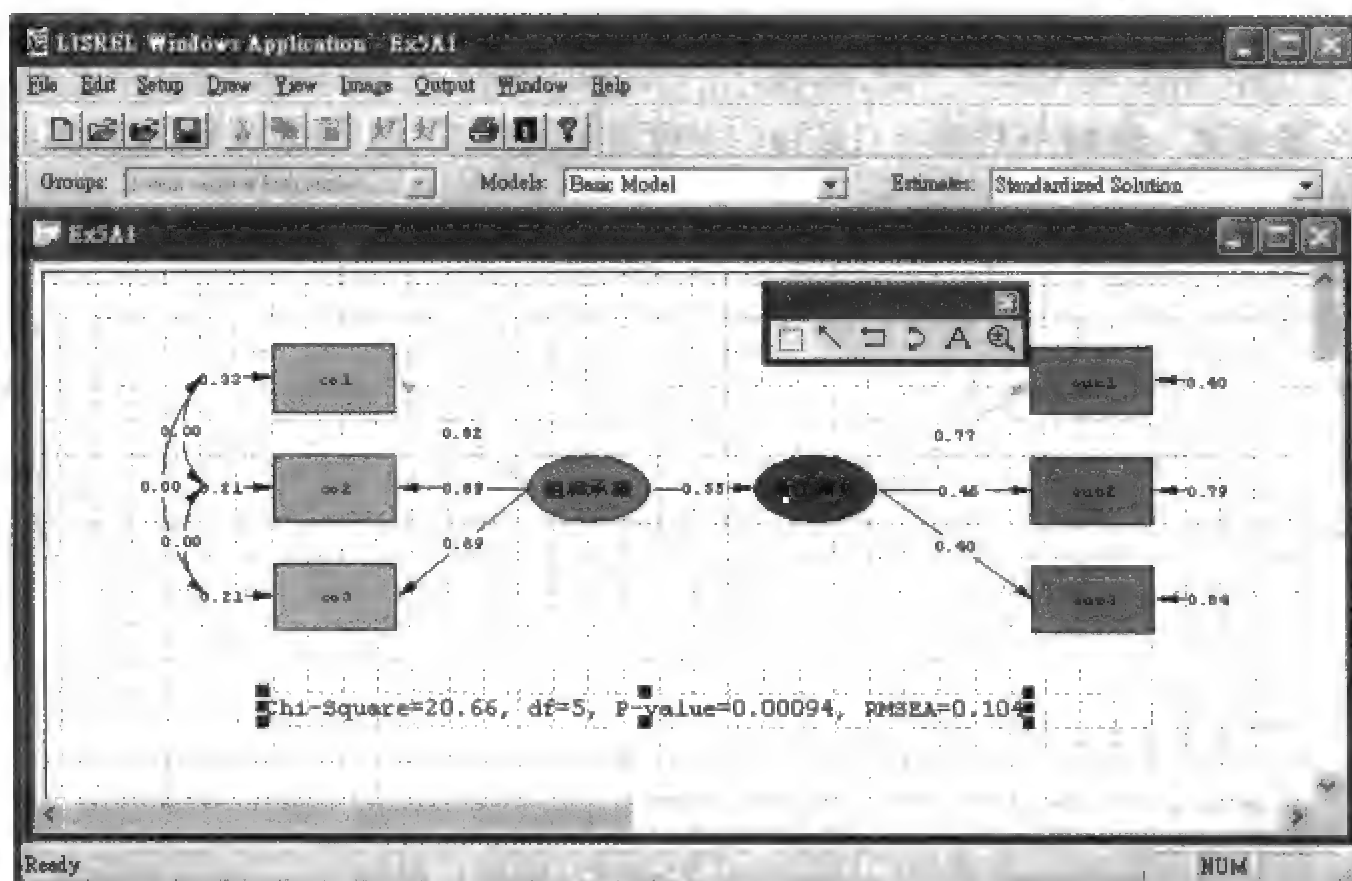


图 1.4 LISREL 程序所产生的标准化解路径图

一般研究在报告 SEM 的分析结果时，多呈现完全标准化解，使每一个参数都具有标准化的性质，例如因素载荷会落入 ± 1 的范围内，如果不是完全标准化解，某些数据可能会出现不寻常数值。但是，在某些时候，研究者仅关心结构模式当中的参数估计，例如路径模型分析，此时标准化解即已足够，结构模式中的参数应会落在合理的范围内，不必针对所有观察变量进行标准化，因为测量模式的参数并不是研究报告的重点。过多的标准化，会使得数据失去原有测量的特性。

值得注意的是，如果研究者输入数据时，是直接输入相关系数矩阵，那么不论

是用 SS 或 SC 指令，得到的标准化解都是完全标准化解，因为相关矩阵的本身就是已经将所有观察变量的数据进行了事前的标准化处理。

四、替代模型的使用

一般来说，SEM 分析很少只处理一个模型的估计。一个研究当中，除了一个基本的模型之外，还可能具有一些替代模型，代表不同的假设或理论模式，经由模型间的比较，以决定何者最能反映实际状态。而替代模型的运用有两种可能：第一是由研究者基于理论观点所提出的假设模型；第二是估计结果所产生的数据所建议的修正模型。前者是一种理论推导演绎的结果，模型的提出与数据无关，因此称为演绎取向（deductive approach）或先验（priori）的替代模型；后者是基于计量数据所得到的替代概念，例如由模型修饰程序所得到的修正模型均属此类，因此称为归纳取向（inductive approach）或事后（post-hoc）的替代模型。

（一）演绎取向的替代模型

最理想的 SEM 模型是基于理论观点所提出的模型，因此，替代模型的提出最好是能够运用演绎取向策略进行。Jöreskog (1993) 指出，任何一个 SEM 模型都可能具有其他的替代模型。因此，SEM 研究应善用替代模型的相互比较，来决定哪一个模型最能够反映实际观察的数据，Jöreskog (1993) 称之为替代模型取向的 SEM 研究。但是，替代模型的提出，是基于理论上或概念上的考虑，因此都是在分析进行之前即已经提出。

演绎取向的替代模型，最重要的价值在于具有先验的理论基础，此外，也正由于模型的建立是在数据分析之前进行，因此模型的设定可以经过详细的计算与安排，不受限于资料的计量特性。在结果的分析上，每一个替代模型的重要性、理论意涵十分清晰明确，从操作面来看，可以免除许多人为操纵的疑虑，减少犯错的可能。

（二）归纳取向的替代模型

如果说演绎取向的替代模型是理论概念的产物，那么归纳取向的替代模型就是现实世界的产物。虽然可能缺乏理论的正当性，但是归纳取向所提出的替代模型却能够反映真实数据的特性，得到的结果最符合真实世界的描述。此类研究 Jöreskog (1993) 称之为模型产生取向的 SEM 研究。

在某些情况下，归纳取向的替代模型分析有其必要性。例如当样本规模很小时，或是某些变量的测量质量较差时，参数估计的稳定性不佳，演绎取向所提出的替代模型之间的比较难以进行。此时，适当的利用参数的估计结果进行修正，可以让参数估计较为顺利。

其次，当研究的性质偏向探索性研究，归纳取向的替代模型可以较演绎取向的替代模型提出更多有建设性的信息。尤其在应用研究领域，例如教育研究、消费调查、管理研究等，理论的引导性不如实际现象与数据具有更高的启发性，此时，从研究数据所反映的各项修饰建议，所累积出来的假设模型更具有解释力。

使用归纳取向的替代模型比较，有两个必须注意的事项。第一，测量残差的相关应谨慎为之，即使修饰指数 (MI index) 显示某些测量残差的相关甚为明显，但是在没有合理的理由或理论观念的支持，测量残差是不宜设定在模型之中的。残差相关的假设，在追踪研究上比较容易看到，因为同一个题目在不同时间点下测量，测量分数除了受到潜在变量的影响，测量残差在不同时间下也应具有一致性，因此，测量残差的相关系数是有其意义的。此外，在多重特质多重方法矩阵 (MTMM matrix) 研究中，同一个测量题目受到不同来源的影响，残差的相关代表的可能是方法效应 (method effects)，因此残差相关也有设定上的理论意义。但是，无论如何，残差相关的设定，除非具有相当的合理性，都不应轻易为之。

第二，残差与其他变量的相关，也不应随意假设其存在。因为测量残差反映了测量误差的存在，而且是具有随机误差的特性；所以，随机变异与其他变量的关系的假设，即使具有统计上的意义，但是一旦被设定在模型之中，整个模型即需要讨论测量误差非随机性的问题，增添了 SEM 分析与解释的复杂度。

一般而言，模型的修饰必须在样本数很大的情况下才比较安全。一般规模的 SEM

研究,例如 100 到 400 个样本,执行模型修饰都有相当的风险,因为如果换一个相同规模的样本重复 SEM 估计,所得到的修饰建议可能是不一样的。此时,随着样本的差异,研究者所得到的结论就会有所不同。这一现象对于强调理论合理性的 SEM 分析是最大的讽刺。

第四节 结 语

综观 SEM 近年来的发展,可以发现 SEM 已经是个高度发展的成熟学科。不仅在学理上有相当程度的进展,在社群活动上也相当的热络。学术活动基本上是由研究人员、学者专家与实务工作者所汇聚的科学社群与具体的研究议题及成果所组成。当一个研究领域聚集了足够的资源,累积了足够的理论论述与研究发现,汇集了相当的共识之后,方能促成并支持专门期刊的诞生与专属社群的动力。SEM 的发展,正是经历了这一连串的学术过程所获致的一个具体成果。

SEM 的相关研究,除了发表在专属期刊《结构方程模型》(*Structural Equation Modeling*)上,也可以在各重要期刊上发现,例如《心理学评论》(*Psychological Bulletin*)、《心理学方法》(*Psychological Methods*)、《心理计量学》(*Psychometrika*)、《应用心理测量》(*Applied Psychological Measurement*)、《多变量行为研究》(*Multivariate Behavioral Research*)、《教育与心理测量》(*Educational and Psychological Measurement*)、《教育测量的议题与实务》(*Educational Measurement: Issues and Practices*)、《社会学方法与研究》(*Sociological Methods and Research*)、《社会学方法论》(*Sociological Methodology*)、《英国数学与统计心理学期刊》(*British Journal of Mathematical and Statistical Psychology*)等,国内有关 SEM 的研究报告则散见在各期刊中。

近年来有关 SEM 在基本原理的重要发展,除了可以应用到非正态化数据的分析,包括二分变量、顺序或类别变量(Browne, 1984, Muthén, 1984),对于特定的问题,例如非随机性遗漏值的处理,Allison (1987)以及 Muthén、Kaplan 与 Hollis (1987)也提出了应对策略。此外,应用在多层次的抽样程序的 SEM 分析(multilevel SEM)以及追踪研究的 SEM 分析,也开始受到研究者的重视与广泛讨论(参见

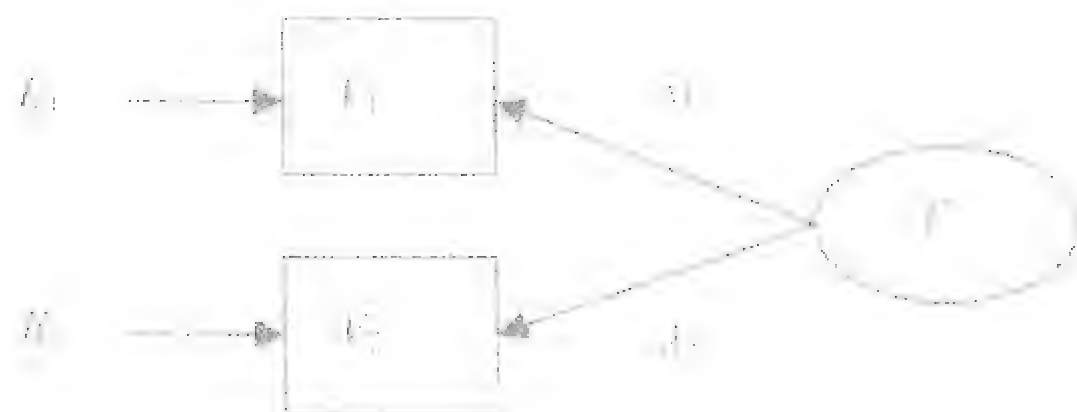
Little, Schnabel, & Baumert, 2000)。一些原创性问题仍不时出现在期刊上，并引起研究者的重视。

结构方程模型虽然有种种的优点，但却面临与其他高等统计技术相类似的命运。不少人对于统计学具有相当的排斥感，使得大多数的研究者未能触及 SEM 之前即已对其抱持敬而远之的心理，加上早期的 SEM 语言过于艰涩，庞杂棘手的各种数学矩阵甚至令熟手亦感头疼，益发凸显了 SEM 学习理解的困难。



第二章

结构方程模型的组成



基本上,结构方程模型是由一系列的变量与参数所组成,联结变量与参数的工作则由数学方程式担任。由于变量与参数众多、关系复杂,因此 SEM 在发展之初的 LISREL 模型即利用数学矩阵整合这些变量与参数的关系。有关 SEM 的基本组成以及所涉及的各个专有名词与数学方程式将在本章中介绍。

事实上,将研究者所关心的变量与参数透过统计模型来描述,进而加以估计分析,即是 SEM 所谓的模型界定(model specification)。透过模型界定,研究者除了可以厘清理论与概念的内容与关系,最重要的是发展成可供结构方程模型进行检验的模型,藉由统计分析的程序,完成所有的评估与统计决策。因此,了解 SEM 模型的组成以及掌握正确有效的模型界定方法,是 SEM 研究的重要环节。

第一节 结构方程模型的变量

一、变量的特性

变量(variable)可以说是统计学中最基本的概念。相对于常数(constant)来说,变量反映了变动与差异,系指某一属性因时地人物不同的内容或是在质或量上的个别差异。基本上,变量可区分成两大类:连续与类别变量。其中连续变量系指利用等距或比率尺度等有特定单位所测量得到的变量。变量中的每一个数值,皆代表强度上意义,又称为量化变量(quantitative variable);相对之下,以当变量数值所代表的意义为质性的概念时,则称为类别变量(categorical variable)或质性变量(qualitative variable)。

类别变量的重点在于分门别类,而非反映强度。因此变量数值仅具相等或不等的数学特性,或是大于与小于的顺序关系,这类变量无法像一般的变量以数学算式计算出有意义的统计量,仅能利用数值的次数(frequency)或百分比(percentage)进行数据的分析与检验。相对之下,连续变量的数值反映了被观察现象或特质的程度大小,因此变量数值可透过加减乘除等数学运算法则获得各种统计量(例如平均数、方差、标准分数等),进行进一步的统计分析。

由于结构方程模型是一套用来分析变量间复杂的共变关系的统计方法，因此结构方程模型的基本组成单位是连续变量，类别变量仅作为辅助或分组讨论的调节变量之用。换言之，在一个 SEM 模型中，凡是被视为变量者，就存在平均数与方差的统计量信息，变量之间的关系即由协方差反映。

二、测量变量与潜在变量

在 SEM 当中，变量有两种基本的形态：测量变量与潜在变量。研究者测量得到的测量变量资料是真正被 SEM 用来分析与计算的基本元素¹；潜在变量则是由测量变量所推估出来的变量。在典型的 SEM 分析中，测量变量的变异系受到某一个或某几个潜在变量影响，因此又称为潜在变量的测量指标（indicators）或外显变量。

在 SEM 的路径图中，测量变量是以长方形来表示的，如图 2.1 的 V_1 与 V_2 。当测量变量是一个无误差的测量之时，我们可以视此变量为一个真实有效的测量，通常人口变量属于此一类型，例如性别、年龄等。另一类的测量变量可能伴随着一定的测量误差，或是反映某种抽象的概念意涵，以 SEM 的术语来说，这些测量变量是受到特定潜在变量的影响，使得测量变量分数呈现高低的变化。

在 SEM 分析的路径图中，潜在变量以椭圆形的符号表示，如图 2.1 的 F 所示。由于潜在变量无法由单一变量反映其抽象内容，必须透过测量变量推估，因此一个潜在变量必须以两个以上的测量变量来估计，称为多元指标原则。不同变量之间的协方差，反映了潜在变量的共同影响。一个 SEM 模型中，测量变量一定存在，但潜在变量不可能单独存在。因为在研究过程当中，潜在变量并不是真实存在的变量，而是由测量变量中所估计出来。

潜在变量的另一个特性是具有测量误差，测量变量的变异无法被共同的潜在变量充分解释的部分称为测量误差，如图 2.1 中的 E_1 与 E_2 。测量误差可以被视为是一个潜在变量，其平均数为零，变异量则可被估计，因此在路径图中有时会以椭圆形符

¹ 测量变量又称为观察变量（observed variable），在本书中这两个名词是代表相同的意思，将交替使用。潜在变量则与因素交替使用。

号表示残差（表示一个潜在变量）（例如AMOS软件的绘图）。

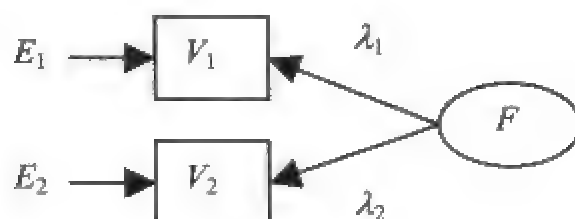


图 2.1 潜在变量与测量变量关系图

三、内生变量与外源变量

沿用过去路径分析的术语，SEM 中的变量可以区分为内生变量与外源变量。所谓内生变量（endogenous variable）是指模型当中，会受到任何一个其他变量影响的变量，也就是路径图中会受到任何一个其他变量以单箭头指涉的变量。外源变量（exogenous variable）则是模型当中不受任何其他变量影响但影响他人的变量，也就是路径图中会指向任何一个其他变量，但不被任何变量以单箭头指涉的变量。

SEM 中，除了以内生及外源变量来区分，如果加上前述测量变量与潜在变量之分，SEM 中的变量可以区分为内生测量变量、外源测量变量、内生潜在变量与外源潜在变量四种类型。当一个潜在变量作为内生变量时，称为内生潜在变量（以希腊字母 η 表示，念作 eta），它所影响的测量变量则称为内生测量变量（以 y 表示）；相对的，当一个潜在变量作为外源变量时，称为外源潜在变量（以希腊字母 ξ 表示，念作 ksi），它所影响的测量变量则称为外源测量变量（以 x 表示）。

若以传统的自变量（independent variable）与因变量（dependent variable）的关系来看，外源变量因为不受他人影响，所以必为自变量，而内生变量多作因变量之用，但也可能作为影响他人的自变量。在回归分析中，因变量视为被解释变量（或称为效标变量），是一项研究主要关心的焦点所在，因此 SEM 中的内生变量可以说是模型中的重要关键变量。



图 2.2 内生变量作为中介变量的关系图

当内生变量同时作因变量与自变量之时，如图 2.2 的 V_2 与 V_3 ，表示该变量不仅被他人影响，进而还可能对其他变量产生作用，因此即成为一个中介变量 (mediator)。由此可知，内生变量之间的关系与本身的参数估计是 SEM 分析相当重要的部分。

内生变量的一个重要性质是具有残差，因为内生变量的变异量不一定能够被模型当中的其他变量所完全解释，其他变量解释内生变量的不足之处，即为残差。对于测量变量而言，其变异量无法被完全解释的残差部分，称为测量残差或独特量 (uniqueness)，如果内生测量变量存在于测量模式当中作为潜在因素的测量指标，残差可以视为是测量误差。

对于潜在变量而言，内生潜在变量的变异量已经排除了测量误差的影响，因此无法被完全解释的残差部分不能视为测量误差，而是其他变量无法解释的独特变异量。这个部分变异量所反映的是模型无法有效解释内生潜在变量的部分，也就是过去回归分析的 $1-R^2$ ，解释因素必定来自于模型之外的因素，因此特别给予一个名词称为 SEM 模型的干扰项 (disturbance)。由于它是无法被其他变量有效解释的变异量，因此也可以称之为解释或预测残差。

第二节 结构方程模型的参数

一、参数的概念

在统计学当中，参数 (parameter) 是一个相当重要但是却很抽象的名词。参数的概念与推论统计有关，带有“未知”与“估计”两个基本特质。当研究者想要从样本

的数据（已知）对于母群体特性（未知）进行了解之时，例如我们想要了解现在学生拥有零用钱的情况，我们可以从一群学生（样本）的每月平均零用钱去推估其他所有学生拥有零用钱的情况（母群体），此时每月平均零用钱称为样本统计量（statistic），所有的学生每月零用钱平均数是我们想要估计得知的参数（母数）。值得注意的是，从统计的观点来看，参数所指的是一个未知而需要进行推估的量数，是一个计量的概念，而非母群体的本身，有时候参数或母数会被误为母体或总体的代名词，是一个不正确的做法。

基本上，SEM 所使用的参数概念与其他统计方法并无差异，但是所包含的类型却较为繁多。在一般的回归分析当中，代表各预测变量对于效标变量预测力的 β 系数即是回归分析的参数；方差分析的主要效应与交互效应也是参数；因素分析当中的因素载荷也是参数的一种。而在结构方程模型当中，可能包括上述各种参数的估计。

研究者所关心的是潜在变量的抽取或是变量之间关系的探讨，因此参数多半与潜在变量的本身有关（例如总体的平均数或方差的估计），或是反映变量之间关系的参数（例如因素载荷、路径系数或协方差的估计）。研究者所拥有的只是样本观察数据的相关矩阵或协方差结构，再利用一些 SEM 软件，可以估计出这些参数数值并进行显著性的假设检验。

二、自由、固定与限定参数

除了参数的性质之外，在 SEM 当中，参数就其估计的需要与否，分成三种类型：自由参数（free parameter）、固定参数（fixed parameter）与限定参数（constrained parameter）。在 SEM 模型中，联结变量与变量关系的关键就是参数，参数的强弱大小必须透过统计程序加以估计，因此都是自由参数。当 SEM 模型中需要估计的自由参数越少时，称为越简效（亦即既简单又有效率）的模型，自由度越大。

其次，SEM 模型当中不被估计的参数，将被设定为 0，因此也被视为是固定参数（固定为 0）。如果因为某些理由被设定为常数（通常是 1）而不被估计者，亦被称为固定参数。

限定参数的使用，多半与多样本间的比较有关，例如某一个参数在甲样本与乙样本间被设定为等值 (equivalent)，此时 SEM 对于这两个参数仅进行一次的估计，是为限定参数。限制参数通常是研究者基于特殊的需要而提出，因此多与特定的理论或逻辑推导有关。而从概念上来看，限定参数介于自由参数与固定参数之间，可以视为半自由参数。但是由于限定参数的数据仍然是由估计得出，因此限定参数与自由参数两者被视为是模型当中必须进行估计的参数。

三、直接与非直接关系

在SEM模型中，参数所联结的变量关系有直接关系 (directional relationship) 与非直接关系 (non-directional relationship) 两种类型。直接关系表示参数带有方向性，变量之间具有假设性的线性因果或预测关系，在路径图中以单向箭头 (\rightarrow) 来表示。非直接关系则表示参数不带有特定方向，亦即变量之间虽然具有关系，但影响方向无法辨认，在SEM路径图中，以带有双箭头的线段 (\leftrightarrow) 或曲线表示。例如图 2.3 当中，潜在变量对于测量变量的关系就是一种直接关系的结构参数，强度以因素载荷表示之。两个潜在变量 (F_1 与 F_2) 关系的直接效应参数则是回归系数，用以说明两者间的影响路径。

四、模型参数与方程式

一个完整的 SEM 模型包括测量模型 (measurement model) 与结构模型 (structure model) 两部分。前者系指实际测量变量与潜在特质的相互关系，后者则说明潜在变量之间的关系，构成这两种模型的参数分别称为测量模型参数与结构模型参数，如图 2.3 所示。

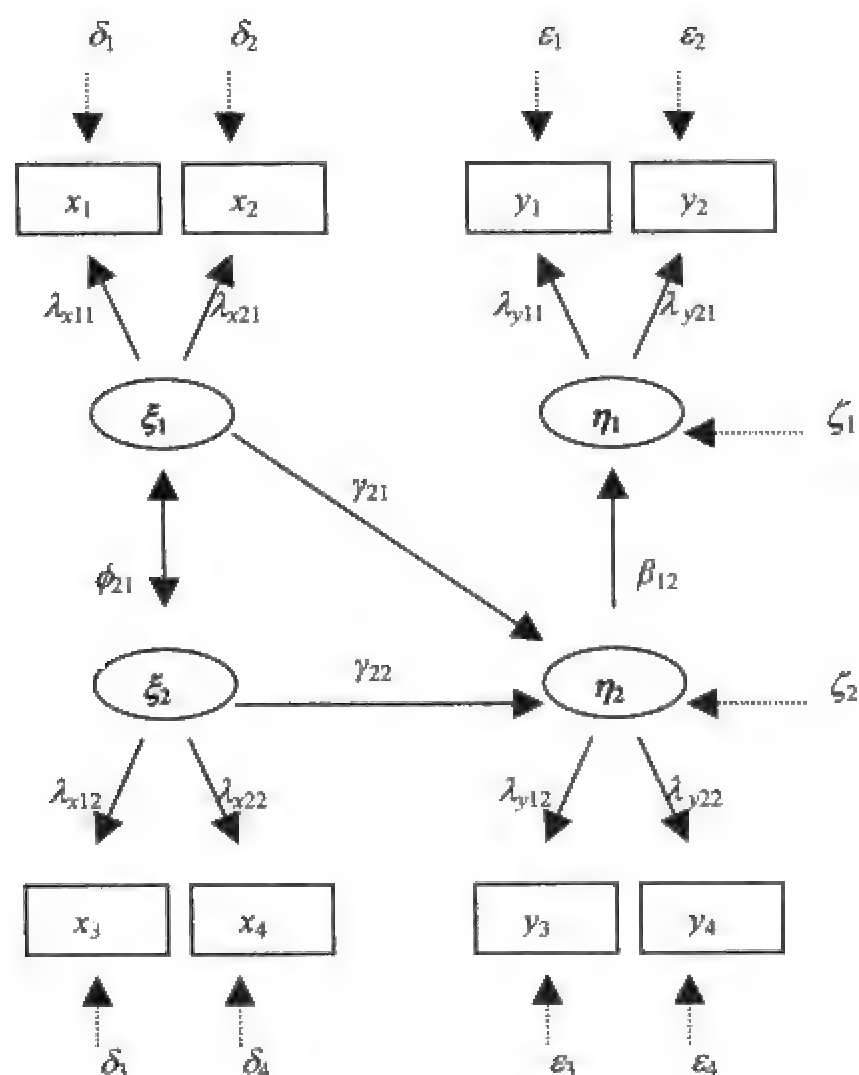


图 2.3 完整 SEM 模型的参数图

图 2.3 中, SEM 模型区分成左侧的外源变量关系与右侧的内生变量关系两部分, 潜在变量与观察变量的关联强度由 lambda (λ) 参数表示, 又称为因素载荷 (factor loading), 所构成的模型即为测量模型。外源与内生潜在变量之间的关系则以 gamma (γ) 参数反映, 内生潜在变量之间的关系由 beta (β) 参数反映, 由 γ 与 β 参数所连结构成的模型即是结构模型。

模型中的 delta (δ) 与 epsilon (ε) 分别表示外源观察变量与内生观察变量被潜在变量解释不完全的测量残差, theta (ζ) 则为内生潜在变量无法被完全解释的估计误差。残差的单箭头表示其方差可以被估计, 而残差之间的共变关系则没有在图 2.3 描述出来, 但可能存在, 可视为自由估计参数来估计之。

上面所提及的各种参数与变量的关系，最后可以利用一般化线性方程式加以描述。其中两个反映测量模式的一般方程式（方程式 2.1 与 2.2）与一个为反映结构模型的一般方程式（方程式 2.3），这三个一般方程式即可构成一个一般化 SEM 模型（general structural equation model），这就是结构方程式（structural equation）一词的涵义：

$$y = \Lambda_y \eta + \varepsilon \quad (2.1)$$

$$x = \Lambda_x \xi + \delta \quad (2.2)$$

$$\eta = B\eta + \Gamma\xi + \zeta \quad (2.3)$$

所谓一般方程式，是指一般性的通式，而不反映特定变量与特定参数的关系。如果针对特定变量与参数的关系，则可利用明确方程式来表现。例如图 2.3 中的结构模式可以下列两个明确方程式来说明参数的关系：

$$\eta_1 = \beta_{12}\eta_2 + \zeta_1 \quad \eta_2 = \gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \zeta_2$$

在 SEM 当中，如果单独使用测量模型，也就是只有测量模型而没有结构模型的回归关系假设时，即为验证性因素分析（confirmatory factor analysis）。因其检测的内容是测量题目的因素结构（factorial structure）与测量误差。进一步而言，单独看待结构模型，其实就是一个传统的路径分析（path analysis）模型，可以多元回归的概念说明变量的因果或预测关系。事实上，如果一个 SEM 模型当中没有任何潜在变量的假设（亦即没有测量模型），只存在有测量变量，并探讨这些测量变量的因果/预测关系时，就与传统的路径模型无异。

值得注意的是，SEM 模型与传统的回归分析不同，SEM 除了可以同时处理多组回归方程式的估计，更重要的是变量关系的处理更具有弹性。在回归分析当中，变量仅区分为自变量与因变量，同时这些变量都是无误差的测量变量。但是 SEM 中，变量的关系除了具有测量关系之外，还可以利用潜在变量进行观察值的残差估计，因此，在 SEM 的模型中，残差的概念远较传统回归分析复杂。其次，在回归分析中，因变量被自变量解释后的残差是被假设与自变量之间的关系是相互独立的。但是在

SEM 分析中, 残差项是允许与变量之间带有关联的, 前提是这些特殊的假设也应具有一定的理论逻辑基础。

在 SEM 中, 测量与结构模型的变量与参数可利用八种主要的矩阵概念来整合。八种矩阵的名称、代号、发音、意义与矩阵状态如图 2.4 所示, 矩阵设定代号见表 2.1。

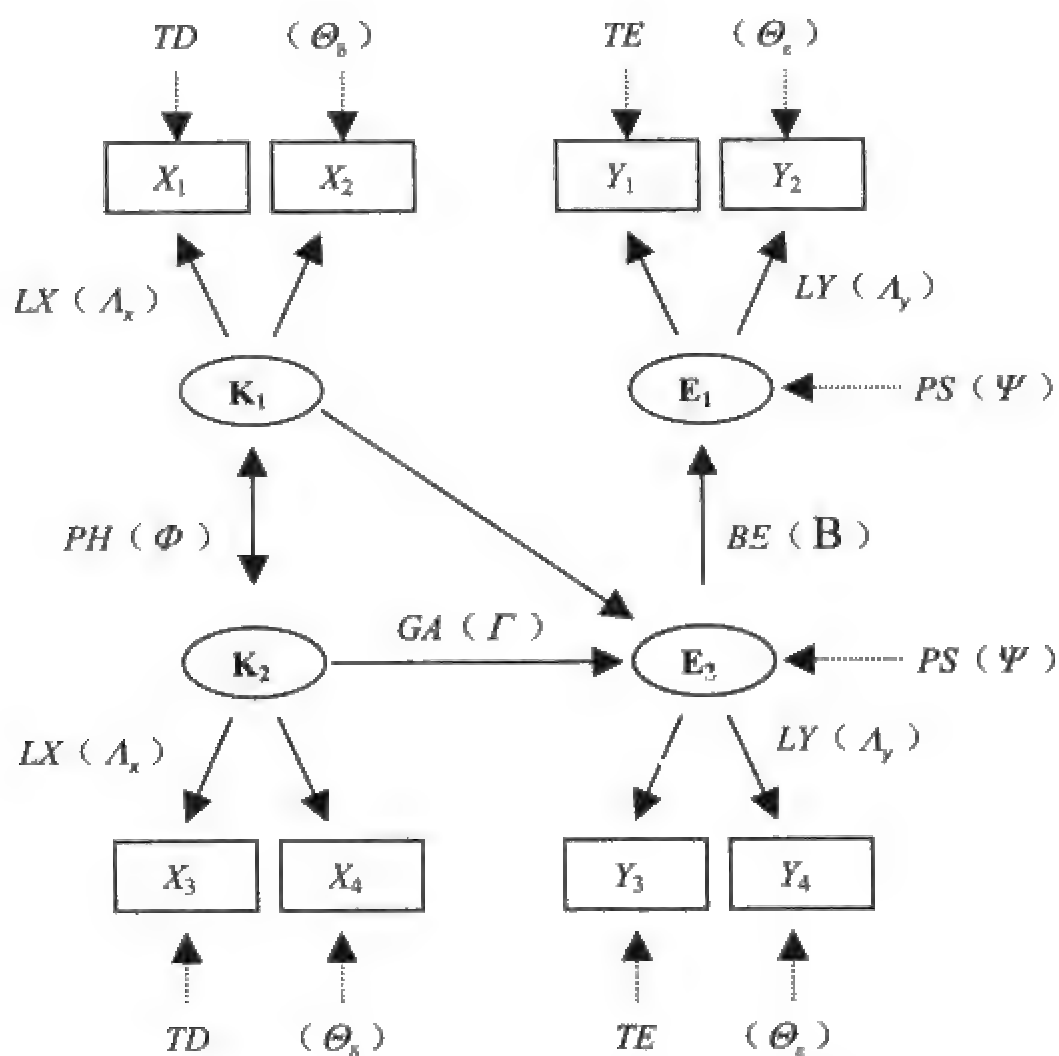


图 2.4 完整 LISREL 模型矩阵的概念图示

表 2.1 SEM 模型的八种矩阵概念列表

符号与发音	缩写	代表意义	<i>mm</i>	<i>mf</i>	order
<u>结构模型矩阵</u>					
B beta	BE	内生潜在变量被内生潜在变量解释之回归矩阵 (E 到 E 的回归系数)	ZE	FI	$NE \times NE$
Γ gamma	GA	内生潜在变量被外源潜在变量解释之回归矩阵 (E 到 K 的回归系数)	FU	FR	$NE \times NK$
<u>测量模型矩阵</u>					
Λ_x lambda x	LX	外源观察变量被外源潜在变量解释之回归矩阵 (K 到 X 的因素载荷)	FU	FI	$NY \times NE$
Λ_y lambda y	LY	内生观察变量被内生潜在变量解释之回归矩阵 (E 到 Y 的因素载荷)	FU	FI	$NX \times NK$
Φ phi	PI	外源潜在变量协方差矩阵 (K 到 K 的因素共变)	SY	FR	$NK \times NK$
<u>残差矩阵</u>					
Ψ psi	PS	内生潜在变量被外源潜在变量解释之误差项协方差矩阵 (解释残差)	SY	FR	$NE \times NE$
Θ_δ theta-delta	TD	外源观察变量被外源潜在变量解释之误差项协方差矩阵 (X 变量残差)	DI	FR	$NX \times NX$
Θ_ϵ theta-epsilon	TE	内生观察变量被内生潜在变量解释之误差项协方差矩阵 (Y 变量残差)	DI	FR	$NY \times NY$

各矩阵当中的向量元素与顺序由观察变量数目 (NX 与 NY) 与潜在变量数目 (NK 与 NE) 决定。其中 NE 为内生潜在变量数目 (m 个)、 NK 为外源潜在变量数目 (n 个)、 NY 为内生观察变量数目 (p 个)、 NX 为外源观察变量数目 (q 个)。

早期以 LISREL 软件进行分析时, 皆以语法 (程序语言) 来设定模型与参数状态, 因此矩阵的特性与关系需要加以设定。在未经宣告时, 自动设定为预设的形式 (model format; *mf*) 与状态 (model mode; *mm*)。传统上的 LISREL 矩阵预设代号列于表 2.2 中, 以*标示。读者亦可自行更改矩阵设定为其他形式与状态, 可供替代的设定代号列于表 2.2 中, 以+标示。

有关矩阵的数值，*ZE* 为 0 矩阵（矩阵中所有数据为 0），*ID* 为相等矩阵（identity matrix，矩阵中除了对角线为 1，其余均为 0），*IZ* 为部分相等与 0 矩阵（partitioned identity and zero matrix），*ZI* 为部分 0 与相等矩阵（partitioned zero and identity matrix）。

在矩阵的形态上，*DI* 为对角矩阵（diagonal matrix），只有对角线的向量被处理，*SD* 为下对角矩阵（sub-diagonal matrix），只有对角线下三角区域的向量被处理，用于回归系数（*B*）矩阵，*SY* 为对称矩阵（symmetric matrix）（包含对角线与下三角区域的向量被处理），但对角线数字不为 1 时为协方差矩阵；*ST* 为对称矩阵且对角线数字设定为 1，用于相关（ Φ ）矩阵；*FU* 为正方或长方形之非对称完全矩阵（full matrix）。至于矩阵参数的估计状态的表示，*FI* 表示固定参数（fixed parameters），*FR* 表示自由估计参数（free parameters）。

表 2.2 LISREL 各矩阵原始设定与替代设定代码对照表

设定代码	矩阵缩写	<i>B</i> <i>BE</i>	Γ <i>GA</i>	Λ_x <i>LX</i>	Λ_y <i>LY</i>	Φ <i>PH</i>	Ψ <i>PS</i>	Θ_ε <i>TE</i>	Θ_δ <i>TD</i>	$\Theta_{\delta\varepsilon}$ <i>TH</i>
Matrix Mode	<i>mm</i>									
Fixed	<i>FI</i>	*	+	*	*	+	+	+	+	*
Free	<i>FR</i>	+	*	+	+	*	*	*	*	
Matrix Form	<i>mf</i>									
Zero(0)	<i>ZE</i>	*					+	+	+	+
Identity(I)	<i>ID</i>		+	+	+	+				
Identity,Zero	<i>IZ</i>		+	+	+					
Zero,Identity	<i>ZI</i>		+	+	+					
Diagonal	<i>DI</i>		+	+	+	+	*	*	*	+
Symmetric	<i>SY</i>					*	+	+	+	+
Subdiagonal	<i>SD</i>	+								
Standardized symmetric	<i>ST</i>					+				
Full	<i>FU</i>	+	*	*	*					

*表示原始设定，+表示可容许。修改自 Jöreskog & Sörbom(1996, p.67)。

第三节 模型界定

一、模型界定的概念

结构方程模型分析的第一个具体步骤就是发展假设模型。透过本章的统计符号与观念所建立的结构方程式，可以协助研究者进行分析与决策。而模型界定的内容可以从两个方面获得。第一是理论的基础或研究者个人的先备知识与经验，第二是 SEM 分析工具的技术语言与方法要求。首先，当研究者面对于所关心的研究问题之时，除了基于自己的知识基础与研究兴趣推演出值得探讨的研究命题之外，还需对于理论文献详加检阅，以建立严谨的科学假设或提出有待验证的理论模型。

进一步，如果研究者决定使用 SEM 检验他所提出的假设与理论模型，则必须配合 SEM 技术语言的规范与各项操作要求，将研究者所提出的假设与理论模型转换成 SEM 模式。在此同时也必需考虑 SEM 分析当中所可能涉及的各种统计概念的基本原理与要求（例如因素分析、路径分析、潜在变量的设定、平均数的估计等），一并纳入 SEM 模型界定之中。这个由理论发展到技术性模型建立的一整套程序，称为模型界定。

在模型界定的过程中，除了依循 SEM 原理设定模式之外，另外一个非常重要的技术问题就是必须让 SEM 模型具有技术上的可识别性，使 SEM 的各项数学估计程序与统计决策过程可以顺利进行。结构方程模型所设定的理论模式是基于研究者的研究需求所提出的。但是，模式的分析与最终统计决策的获得，必须利用实际搜集得到的数据，转换成为 SEM 分析的数据格式之后，利用分析软件来进行分析工作。只有在假设模型符合统计分析 with 软件执行的要求，也就是在能够被有效识别的情况下，SEM 分析才能顺利进行。此时，一个模型可以被分析工具有效识别的程度称为模型识别度 (identifiability)，有关议题将在下一章介绍。

二、模型界定的简效原则

共变结构关系的分析可以说是 SEM 最主要的核心概念。一个 SEM 模型往往涉及数十个变量的共变关系的探讨，这些变量可以组合成无数种不同的假设模型。当 SEM 分析牵涉到如此复杂的变量关系的探究时，一个重要的基本原则是如何将这些共变关系以最符合理论意义且最简单扼要的方式来加以界定，并使之能够最符合实际观测到的数据结构。此一原则称为简效原则（principle of parsimony）。

简效原则反映了 SEM 分析的一个主要限制，就是研究者无法精确说明究竟一组变量的共变结构关系应该以何种模型来表示为最好。因为同样一组变量的组合有无限多种可能，而不同的关系模式可能都代表了特定的理论意义，因此，如果研究者可以用一个比较单纯简单的模型解释较多的实际观察数据的变化，那么以这个模型来说明变量间的真实关系，比较不会得到错误的结论，避免犯下第一类型错误。换句话说，SEM 分析对于理论模型的检验，只能说明一个模型不至于有错，但是却很难证明它是真正正确的模型。

三、等值模型问题

从 SEM 出现以来，有一个常被提出讨论的议题，是所谓等值模型的问题（Marcoulides & Schumacker, 1996; Raykov & Marcoulides, 2001）。所谓等值模型（equivalent models）是指不同的 SEM 模型具有相同的模型拟合度，也就模型拟合指数（例如 χ^2 与 p value）统计量在不同的模型间具有相等的数值。等值模型的出现，使得统计上“相等”的模型，却有不同的解释方法。如此一来，将使得 SEM 的研究产生相当大的困扰。不仅在概念层次上无法有效地厘清变量的关系与模型的优劣，在 SEM 的操作层次，也使得模型比较无法进行。

另一方面来说，即使等值模型的现象没有发生在某一个研究中，SEM 的操作者也必须明了潜在的等值模型现象是可能存在的。换句话说，任何模型的提出，都有可能发生模型界定方式不同但是模型的拟合度却相同的情形。研究者对于 SEM 分析

结果的解释，虽然是就他所提出的模型以及所获得数据来解释，但是同样的一套数据，却也可能以不同的方式来诠释其意义。有许多研究已经证实，对于特定的 SEM 模型，操作者可以在没有任何理论基础的情况下，利用 SEM 模型界定策略提出具有相同拟合度的等值模型（例如，Hershberger, 1994; Lee & Hershberger, 1990; Stelzl, 1986; Raykov & Penev, 1999）。

等值模型的现象可以用一个最简单的回归分析的概念来说明。在相关或协方差的概念中， X 与 Y 两个变量关系的统计量并没有特定的方向性，两个变量可以下列关系式描述：

$$X \leftrightarrow Y \quad (2.4)$$

上述 X 与 Y 的关系如果改以回归的方式来解释，则产生了两种情形：

$$X \rightarrow Y \quad (2.5)$$

$$X \leftarrow Y \quad (2.6)$$

前者 (2.5) 为以 X 变量去解释 Y 变量的变异量，后者 (2.6) 是以 Y 变量去解释 X 变量的变异量。两者具有不同的回归方程式，各自计算出一组回归系数。但是如果将回归系数进行标准化，我们会发现两个方程式具有相同的 Beta 系数，且等于相关系数。三个模型其实是同一个模型，但却有不同的理论意义，造成 SEM 模型比较决策的问题。

对于等值模型现象在 SEM 分析中的影响，文献中已有相当数量的讨论。目前最常被引用的解决方案称为前导理论 (priori theory) 策略 (MacCallum, Wegener, Uchino, & Fabrigar, 1993)，它仍然非常原始地从模型发展的初始，以特定理论基础或前导研究发现作为模型界定的依据，而其他各种等值模型则视为不合理但可能存在的模型。如果没有必要，一般皆把等值模型排除在研究讨论之外，以减少不必要的困扰。

另一个普遍的做法是提出几个具有特殊理论意涵的代表性对等模型，利用 SEM

估计获得各参数的估计数，透过参数估计的比较与分析，分析哪一个对等模型具有较高的解释力。此一策略我们可以称之为参数竞争比较（parameter comparison）策略。由于对等模型的卡方值相等，因此所有的拟合指数都失去了比较的意义，仅可逐一检查参数估计结果来检视模型的优劣。

虽然等值模型在单一模型的评估上有其困难，但是每一个等值模型可以继续发展不同的嵌套模型以进行竞争比较。因此，Kline（1998）即建议可以利用代表性的等值模型，继续进行等值模型各自的进一步的竞争比较，以检验哪一个等值模型具有最佳的解释性。然而，这些做法最终的决策，仍在于何者最符合理论上的解释或最具有逻辑推导的正当性，一再显示出理论的考虑在 SEM 分析中的重要性。

第三章

参数估计与识别问题



在结构方程模型的分析过程中，最核心的计量程序就是参数估计（parameter estimation）。由于完全交由计算机处理，所以这也是 SEM 使用者最难以一探究竟的部分，可以说是 SEM 分析过程中最大的黑箱。

参数估计的进行建立在特定的统计计量原理之上。因此，用于 SEM 分析的测量数据，除了必须符合一般性的统计假设之外，还必须符合属于 SEM 特殊的要求。例如一个 SEM 模型必须具有统计与方法上的可识别性，方能使各项估计程序与统计决策过程可以顺利进行，称为模型识别（model identification）议题。本章的主要目的，即在介绍 SEM 的参数估计方法与识别问题，以利使用者了解 SEM 的运作原理。

第一节 模型识别问题

一、参数数目与数据点（DP）

识别性的问题可以模拟为我们中学时代解联立方程式的过程，例如，对于一个二元一次方程式：

$$X+Y=10 \quad (3.1)$$

如果没有指定 X 或 Y 值或提供第二组方程式的情况下，符合该方程式的条件的 (X, Y) 解有无限多种可能，例如 $(2, 8)$ 、 $(5, 5)$ 都是可能解。此时，就是所谓无法识别或识别不足（underidentified）的状况，也就是数学上无法得到一组特殊解的现象。除非我们提出第二个方程式：

$$X+2Y=20 \quad (3.2)$$

此时，即能够求出一组及唯一一组特殊解 (0, 10)。此种当我们提供足够的条件去求取联立方程式未知变量的解，以 SEM 的术语来说，即是一种可以识别的状况，称为充分识别 (just-identified)。但是，我们若再提出第三个方程式，例如：

$$X+3Y=40 \quad (3.3)$$

此时，两个未知数却有三个方程式，也会造成无特定解的状况，但是此时却可以利用估计的方式，求出符合这三个方程式的最佳解，称为过度识别 (over-identified)。例如以 (0, 10) 代入三式，分别得到 10、20、30，此解在第三个方程式产生 10 的差距 (残差)；若改以 (-1, 11) 代入三式，则可以得到 10、21、32，此解在三个方程式的残差分别为 0、1、8，总和为 9，也就是说，(-1, 11) 是一个较佳解 (better solution)。

结构方程模型的理论模型多是由好几个多元一次联立方程式所组成。已知的条件与未知的参数数目的估算相当复杂，稍有不慎，就会造成模型无法估计 (待解未知数多于方程式数目) 的窘境。因此，在 SEM 当中，有关联立解的求取与估计的方法也就格外受到重视。

传统上，模型识别系以整体模型为计算基础 (Bollen, 1989)，但 McDonald 与 Ho (2002) 指出识别性问题应从三个角度来看：测量模型、结构模型与潜在变量的界定。下面我们分别就 SEM 模型的不同层次的识别性来说明。

二、整体模型识别性

决定模型识别性的具体步骤首先是计算用以产生共变结构的观测值数目，称为测量数据数 (the numbers of data points, DP)。测量数据数与样本测量变量共变量矩阵当中的协方差与方差数目有关，可利用下式来计算：

$$DP = \frac{(p+q)(p+q+1)}{2} \quad (3.4)$$

其中 $p+q$ 表示测量变量的个数，其中 p 为外源测量变量的数目， q 为内生测量变量的数目。假设今天有 10 个测量变量，总计可以产生 10 个方差与 $C_2^{10}=45$ 个协方差，合计为 55 个数据点， $DP=55$ 。至于被估计的参数数目的计算，则牵涉到研究者所提出模型的各种设定，包括了回归系数、协方差、方差三类参数。这些未知的参数必须能够顺利地估计。

（一） t 法则

Bollen (1989) 利用 DP 数与参数估计数目的比较判断模型的识别性，提出了一个衡量识别性的必要但非充分的识别条件计算法则 t 法则 (t -Rule)， t 值代表模型中的自由估计参数数目。SEM 模型能够被识别，必须符合下列关系式：

$$t \leq \frac{1}{2}(p+q)(p+q+1) = DP \quad (3.5)$$

t 法则的判断原则如下：

1. 当 $t < DP$ 时，为过度识别，好比我们有过多的方程式，但是只需求取少数几个因素解；
2. 当 $t = DP$ 时，为充分识别，好比我们用两个方程式来求二元因素的解；
3. 当 $t > DP$ 时，为识别不足，如同以太少的条件求取过多的因素解。在 SEM 分析中，识别不足的情况将导致无法进行任何参数估计。

在充分识别的情况下，参数估计恰可以导出一组估计协方差矩阵且完全等值于样本观测协方差矩阵，因此又称为饱和模型 (saturated model)。在饱和模型下，估

计模型与实际模型共变结构完全等值,卡方统计量为 0,呈现完美拟合。虽然参数估计结果稳定,也可以检测每一个个别的参数,但是无法评估整体模型的适切性,无法进行模型拟合度的假设检验。只有在过度识别的情况下,SEM 可以利用不同的参数估计方法,对于参数进行优化的估计,也就是从无限组解当中找到最佳解,进而从参数估计的结果当中,得出整体模型拟合度的评估,或进行模型的比较。

在识别不足的情况下,若要使 SEM 分析能够顺利进行,研究者可以将部分参数设定为定值,也就是不予估计,或是将参数进行设限,或直接将参数估计移除在模式之外,使测量数据点能够大于 q ,即可顺利进行 SEM 分析。

(二) 虚无 B 矩阵法则与递归法则

当一个 SEM 模型中没有任何的内生变量去影响其他的内生变量时,也就是没有结构关系的假设时,也就没有任何的结构参数(β 系数)的估计,亦即 B 矩阵为 0。此时 Φ 、 Γ 与 Ψ 等各矩阵未知参数数目为导出矩阵(Σ 矩阵)识别参数数目的函数,因此整个模型可以顺利识别。换句话说,当 B 矩阵为 0 时,模型自动识别,Bollen (1989)称之为虚无 B 矩阵法则(Null B Rule)。Null B Rule 可以说是反映模型识别性的充分条件,当一个 SEM 模型若符合 Null B Rule,则模型的识别性不会成为问题,不必计算识别性数值。

另一个模型识别的充分条件,是模型中当 B 矩阵呈现三角形状态(对称矩阵,所有变量间的结构参数均加以估计),而 Ψ 矩阵呈现对角线状态(表示仅估计干扰项的方差,干扰项的相关不列入估计),此时为一递归模型且为饱和模型。符合此条件的模型会被模型自动识别,称为递归法则(recursive rule)。

在递归模型(recursive model)中,变量的因果关系是单一方向,预测残差项是彼此独立的独立残差模型;非递归模型(nonrecursive model)则允许干扰项具有相关的假设关系而列入估计,变量间亦可能具有回溯关系。当递归模型条件成立时, B 、 Φ 、 Γ 与 Ψ 等各矩阵自动获得识别,因此递归法则显示的也是识别性的充分条件。

整体来说,要判断一个模型是否能够识别,牵涉到模型的界定方式当中各矩阵与各参数的不同状况的影响,初学者非常容易碰上无法识别的尴尬状况。为了解除

识别不足的状况，Hayduk (1987) 建议使用者尽量简化 SEM 模型，如果识别不足的情况持续存在，建议同时将潜在变量的残差设定为 1，或是将信度理想的测量模式的参数设定为 1，或去除不良的变量（结构参数偏低），或许能够获得识别。

三、测量模型的识别性

在 SEM 分析中，测量模型所决定的是整体模型当中的外显变量（测量变量）与潜在变量之间的关系。一个结构方程模型可能包括了多组测量模型，对于每一个测量模型的识别性，也应符合识别性原则，否则可能影响整体模型评估的稳定性。

对于个别测量模型的识别，主要牵涉到潜在变量量尺的设定问题。因为潜在变量是由 SEM 分析所估计出来的，而非实际的观测变量，因此必须给定一个特定的单位尺度。此时，可以将潜在变量的方差设定为 1，也就是将潜在变量强制以标准化方差作为共同单位。另一种做法是将潜在变量所影响的各测量变量中，指定其中一个变量的因素载荷（或回归系数）为 1，也就是将测量变量的单位设定为潜在变量的参照量尺，并使潜在变量的方差得以自由估计。

当潜在变量作为影响别人的外源变量时（外源潜在变量），上述两种方式都可以用来设定测量模型的单位。但是当潜在变量作为被他人决定与影响的内生变量时（内生潜在变量），潜在变量的本身成为估计的主要对象，此时应设定测量变量的因素载荷为 1，使潜在变量的方差可以自由估计。

一个理想的测量模型，应是潜在变量所影响的每一个测量变量的因素载荷越接近 1，且需达到统计的显著性。同时每一组测量模型各自形成一个独立的丛集，彼此间没有假设性的因素负荷假设，也就是每一个测量变量仅受到单一潜在变量的影响。此时测量变量被称为纯化指标（pure indicator），此要求是一个非常严格的模型界定策略。

在严格的模型界定策略下，测量模型的识别性的计算主要涉及测量模型中所设定的每一个潜在因素所影响的测量变量数目多寡。对于只有一个潜在变量的测量模式，至少要有三个测量变量才能满足测量模型的识别性。同时这三个测量变量的因

素载荷都必须不等于 0, 且测量残差之间没有任何相关的假设。

如果测量模型当中有超过一个以上的潜在变量, 每一个潜在变量只要有至少三个测量变量来估计, 而每一个测量变量只用以估计单一一个潜在变量 (或被单一一个潜在变量所影响)、残差之间没有共变假设且潜在变量的方差被自由估计, 此时测量模型可以有效被识别。另外, 如果潜在变量只以两个测量变量来估计, 残差无相关、每一个测量变量只用以估计单一一个潜在变量且没有任何一个潜在变量的共变或方差为 0, 则模型可以维持可识别的状况。

McDonald (1999) 提出一个较为温和的模型假定策略, 称为独立丛集策略 (independent cluster basis)。当潜在因素之间具有相关假设时, 每一个潜在变量至少需要两个纯化指标; 或是当潜在因素之间不具有相关的假设时, 每一个潜在变量需要至少具有三个纯化指标。此概念是一个简单的概念, 可以套用在验证性因素分析的模型识别性计算。McDonald 与 Ho (2002) 的分析指出, 多数 SEM 研究均忽略了测量模型的识别性的报告, 使得读者无从判断测量模型的设定是否恰当, 因此呼吁研究者应主动提供测量模型的识别性信息。

四、结构模型的识别性

对于整个 SEM 模型的识别性, 如果模型当中包含有结构模型 (也就是统合模型的情况), 除了测量模型的识别问题之外, 必须进一步地评估结构模型的识别性, 以确认结构模型当中的参数设定是恰当的。在计算上, 结构模型识别性与测量模型识别性是两个独立的判断历程。结构模型的识别性判定主要牵涉结构模型的结构参数的设定, 也就是潜在变量之间的参数数目无关乎测量模型的参数设定。

结构模型的识别性决定于内生变量之间的关系的假设。如果内生变量彼此之间没有任何预测关系假设 (B 矩阵为 0), 可直接应用前面提及的虚无 B 矩阵法则 (Bollen, 1989), 此时结构模式必然可以识别。如果内生变量彼此之间具有特定的预测关系假设, 此时需检验这些预测关系是否属于递归模式的路径模型。如果该结构模型是递归模式模型, 也就是说没有回溯关系、没有干扰残差的相关假设, 此时结构模型仍然可以识别, 也就

是递归法则的概念。

在非递归模式的路径模型中，潜在内生变量彼此之间具有回溯关系（feedback loop），干扰残差项具有相关的假设，此时结构模型因为涉及过多参数的估计而无法识别，研究者必须使用别的策略使模型能够识别。除此之外，非递归模式的结构模型有两个外加的条件。第一，每一个方程式至少要有潜在内生变量数目减一个变量不属于非递归模型路径；第二，用以计算标准误的讯息矩阵（information matrix）必须可以被完全估计，并可以求出倒置讯息矩阵（inverted information matrix）。如果从 SEM 的分析软件中无法得到一个倒置讯息矩阵，表示第二个原则无法满足，结构模型的识别性不足，在进行 SEM 分析时，将会出现警语，警告基于无法求出倒置讯息矩阵的识别不足问题。

McDonald (1997) 提出了前置原则（precedence rule）与直交法则（orthogonality rule）的概念，是两种较简易的判断原理。所谓前置原则是指具有因果的先后次序的变量中（也就是具有因果关系的那一对变量），干扰项协方差均设定为零，而当所有的内生变量的干扰项之间都不具有相关的假设时，称为直交法则。在这两个法则下，结构模式的识别性是足够的。在一般的递归模型中，直交法则可以作为判断识别性的依循；但是在非递归模型中，直交法则便不适用。¹

五、潜在变量的量尺化与识别性

潜在变量与一般测量变量最大的不同在于其“不可直接测量”的特性，因此潜在变量缺乏一个自然存在的尺度，而必须以人为的手段设定尺度。这里所指的尺度，可以被视为测量使用的量尺或单位，例如身高是一个测量变量，单位是厘米，属于比率尺度。但是对于智力这一个潜在变量，它的单位或尺度就必需由测量者依据智力估计的方法来加以设定。

在 SEM 当中，潜在变量的尺度因为设定方法的不同而不同，具有非决定性（indetermination）。这往往会造成 SEM 参数估计的问题。对于 SEM 模型中的外源潜在

¹有关非递归模型的辨识性问题，可参考Bollen (1989) 的说明。

变量的设定,最常使用的方法是将潜在外源变量的方差设定成一个常数(通常为1.00),也就是将模型当中所有的潜在变量进行标准化,在方差维持一致,如此可以对于其他参数进行比较。对于潜在内生变量,通常是将其中的一个测量变量与潜在变量的因素载荷设定为常数(通常为1.00)。

针对潜在变量进行量尺化(scaling)设定。具体做法有二,也就是可以将潜在变量的方差设定为1,或是指定潜在变量中任何一个变量的因素载荷(或回归系数)为1。但是,当一个SEM模型混合了测量与结构模式时,潜在变量的量尺化便较为复杂。

在兼含测量与结构模式的统合模型中,对于外源潜在变量的设定,可以单纯地将外源潜在变量的方差设定为1来进行量尺化。但是,对于内生潜在变量,由于作为内生变量其变异量是被外源变量解释,并非被估计参数,因此内生潜在变量的量尺化必须特别小心处理,否则将造成外源变量无法顺利估计内生潜在变量方差的识别错误。

内生变量的量尺化有三种常用策略,第一是Browne与Du Toit(1992)所提出,将内生潜在变量的方差设定为1,进行设限的参数估计程序。第二是McDonald、Parker与Ishizuka(1993)的再参数化策略(reparameterization method),也是将内生潜在变量的方差设定为1,但是仅应用于递归模型的设定。第三是取内生潜在变量其中一个因素载荷为固定参数,进行参数估计后,将潜在变量量尺固定后,再进行内生潜在变量的方差估计,称为再量尺法(rescaling method)。

前两种策略分别可由不同的分析软件来设定(例如RAMONA、SEPATH、CALIS、COSAN)。其优点是在完全标准化的参数估计下,标准误能维持不偏,但是结构参数的估计数值存在偏差。第三种再量尺策略应用于LISREL与EQS软件,优点是在完全标准化的参数估计下,可获得正确的参数估计值,但是某些标准误则无法正确估计,使得显著性检验参考性降低。目前大多数的SEM研究均是以LISREL或EQS所提供的再量尺法进行内生潜在变量的量尺化设定,使得这些研究能够获得正确的标准化解,但是无法获得充分的标准误与显著性检验数据。

McDonald与Ho(2002)指出,虽然标准化估计解的报告是证明理论模型各参数意义的重要数据;因此强调精确的标准化解是一个正确的做法,但是并不意味着

内生潜在变量中，参数估计的显著性检验检查可以被忽略。因此，搭配不同的检验程序，使参数估计的显著性检验能够被正确的报告，是未来研究可以努力的目标。但是至少在目前，LISREL 与 EQS 等常用软件仍未能积极处理此一问题。

第二节 参数估计

一、相关与共变

相关 (correlation) 是反映两个变量线性关系强度的统计概念。两个连续变量的关联情形，除了用散点图的方式来表达，可以利用数学模型来呈现其计量特性，也即是建立一个用以描述相关情形的量数，即相关系数 (coefficient of correlation) 来表示线性关系的强度。

线性关系的定义是可以以一条最具代表性的直线表示两个变量的关系，并以直线方程式中的斜率表现变动关系的计量关系。但是，斜率并不足以说明两个变量观察值的分布情形。若要相关系数能反映两个变量的配对观察值的分布，其运算必须考虑到两个变量各自的集中与分散状况，以及配对分数的集中与分散状况，将所有观察值的分布情形纳入考虑，以协方差的概念进行。而结构方程模型即是用来处理多个变量之间复杂共变结构的统计技术。协方差的公式如下：

$$Cov(X, Y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N - 1} \quad (3.6)$$

协方差的正负号代表两变量是正向或负向关系。例如，若要得到一个正值的协方差，两个离均差必须同时为正值或同时为负值，也就是两个变量需同时在平均数的左侧或右侧，表示两个变量有同方向的变动关系，或正向关联。相反的，要得到一个负的协方差时，两个离均差必须同时一为正值、一为负值，也就是两个变量有

反方向的变动关系，或负向关联。

值得注意的是，协方差的数值会因为两个变量的不同单位，而没有一定的范围，因此协方差的大小可以反映两个变量共同变化时的原始量的变动，但两个协方差数值无法直接用于比较。若将公式 3.6 除以两个变量的标准差，协方差即成为相关系数，也就是去除单位的标准化关联系数。公式如下：

$$r = \frac{Cov(X,Y)}{s_x s_y} \quad (3.7)$$

相关系数作为一个标准化的协方差，其目的在反映两个变量关联强度，当系数值越高，表示线性关系越强，系数值越低（绝对值越接近 0），表示线性关系越弱。结构方程模型的分析过程，主要是探讨两两变量的变化关系的结构特性，因此，适当的数据格式是输入协方差数据，协方差的优点是带有各变量的原始计量特征。如果是输入相关系数，因为标准化过程已经消除了变量的原始分布特征，所以必须另外输入平均数与标准差，在分析时才可获得测量变量的完整性质。

二、SEM 中的共变推导

SEM 对于参数的估计，主要与方差及共变结构的导出过程有关。Hays (1994) 指出四种与方差（协方差）计算有关的定理，说明如下：

定理一：某一个变量与自己的共变即等于该变量的方差。亦即：

$$Cov(X,X)=Var(X)$$

定理二：经过线性整合后的变量的协方差为：

$$Cov(aX+bY,cZ+dU)=acCov(X,Z)+adCov(X,U)+bcCov(Y,Z)+bdCov(Y,U)$$

定理三：经过线性整合后的变量的方差为：

$$Var(aX+bY)=Cov(aX+bY,aX+bY)=a^2Cov(X,X)+b^2Cov(Y,Y)+2abCov(X,Y)$$

定理四：独立的两个变量的线性整合后的方差为：

$$Var(aX+bY)=a^2Cov(X,X)+b^2Cov(Y,Y)$$

这四个定理在概念上非常简单易懂，也是 SEM 利用共变结构进行分析的主要原理，尤其是第三、四个定理所显示的，当两个变量之间没有关联时，也就是 $Cov(X,Y)$ 等于 0，两个变量线性整合后的方差等于两个变量方差的和，而且此一关系不受合并参数 (a, b, c, d) 的影响。当一个 SEM 模型当中有许多变量进行共变结构分析时，变量间独立性与相依性的差别，就在于第三与第四法则应用上的差异。

三、方差与协方差导出矩阵

现以图 3.1 中的结构方程模型分析为例。对于两个观察变量 V_1 与 V_2 ，其数学方程式为：

$$V_1 = \lambda_1 F_1 + E_1$$

$$V_2 = \lambda_2 F_1 + E_2$$

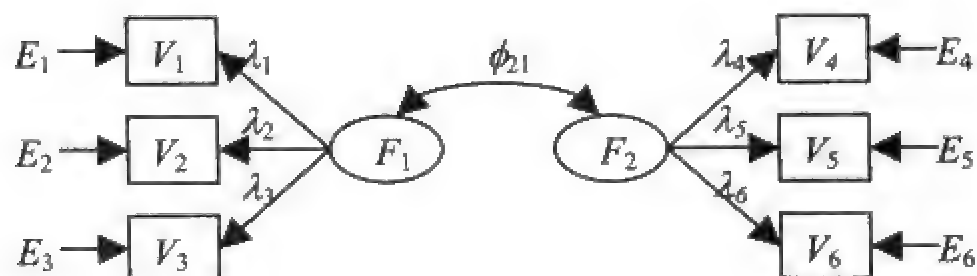


图 3.1 两个具有相关的潜在变量的 CFA 图

由定理二，计算两个观察变量的协方差为：

$$\begin{aligned}
 \text{Cov}(V_1, V_2) &= \text{Cov}(\lambda_1 F_1 + E_1, \lambda_2 F_1 + E_2) \\
 &= \lambda_1 \lambda_2 \text{Cov}(F_1, F_1) + \lambda_1 \text{Cov}(F_1, E_2) + \lambda_2 \text{Cov}(E_1, F_1) + \text{Cov}(E_1, E_2) \\
 &= \lambda_1 \lambda_2 \text{Cov}(F_1, F_1) = \lambda_1 \lambda_2 \text{Var}(F_1, F_1) = \lambda_1 \lambda_2
 \end{aligned}$$

要得到上式的结果，必须符合三个条件：第一，两个误差项的共变为 0；第二，误差项与潜在变量的共变为 0；第三，潜在变量 F_1 的方差为 1。也就是说，符合前述条件时，受到同一个潜在变量影响的两个观察变量，其协方差等于潜在变量所属的两个观察变量的因素载荷乘积。

进一步的，对于 V_1 与 V_4 这两个不同潜在变量的观察变量，其协方差计算式为：

$$\begin{aligned}
 \text{Cov}(V_1, V_4) &= \text{Cov}(\lambda_1 F_1 + E_1, \lambda_4 F_2 + E_4) \\
 &= \lambda_1 \lambda_4 \text{Cov}(F_1, F_2) + \lambda_1 \text{Cov}(F_1, E_4) + \lambda_4 \text{Cov}(E_1, F_2) + \text{Cov}(E_1, E_4) \\
 &= \lambda_1 \lambda_4 \text{Cov}(F_1, F_2) = \lambda_1 \lambda_4 \phi_{21}
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(V_1) &= \text{Cov}(\lambda_1 F_1 + E_1, \lambda_1 F_1 + E_1) \\
 &= \lambda_1^2 \text{Cov}(F_1, F_1) + \lambda_1 \text{Cov}(F_1, E_1) + \lambda_1 \text{Cov}(E_1, F_1) + \text{Cov}(E_1, E_1) \\
 &= \lambda_1^2 \text{Var}(F_1) + \text{Var}(E_1) = \lambda_1^2 + \theta_1
 \end{aligned}$$

也就是说，观察变量的方差，等于各观察变量的因素载荷的平方加上误差项的方差 (θ_1)。

利用上述各式的概念，可以逐一计算出图 3.1 当中的六个观察变量的方差与配对协方差，进而产生一个由参数所导出的方差与协方差矩阵 (Σ)，内容如下：

$$\Sigma = \begin{bmatrix} \lambda_1^2 + \theta_1 & & & & & \\ \lambda_1\lambda_2 & \lambda_2^2 + \theta_2 & & & & \\ \lambda_1\lambda_3 & \lambda_2\lambda_3 & \lambda_3^2 + \theta_3 & & & \\ \lambda_1\lambda_4\phi_{21} & \lambda_2\lambda_4\phi_{21} & \lambda_3\lambda_4\phi_{21} & \lambda_4^2 + \theta_4 & & \\ \lambda_1\lambda_5\phi_{21} & \lambda_2\lambda_5\phi_{21} & \lambda_3\lambda_5\phi_{21} & \lambda_4\lambda_5 & \lambda_5^2 + \theta_5 & \\ \lambda_1\lambda_6\phi_{21} & \lambda_2\lambda_6\phi_{21} & \lambda_3\lambda_6\phi_{21} & \lambda_4\lambda_6 & \lambda_5\lambda_6 & \lambda_6^2 + \theta_6 \end{bmatrix}$$

此一由各参数所导出的 Σ 矩阵，称为导出矩阵（reproduced matrix）。也就是说，矩阵中的每一个量数，是由 SEM 的假设模型与参数估计所推导得出，而非实际测量观测到的数值。

对于 Σ 矩阵中的每一个元素，具有一个对应的实际观测值。也就是说，由样本测量得到的 6 个观察变量，它们的方差与协方差也可以利用一个 6×6 的矩阵来表示，称之为 S 矩阵。矩阵中的每一个数值都是具体的实际观测值。将两个矩阵的数值相减，可以得到一个残差矩阵，用来评估每一个量数的拟合情形。所有残差整合之后，即可以用来评估理论模型与实际模型的拟合度。

第三节 参数估计策略

前面一节指出 SEM 进行分析的主要原理是方差与协方差。利用共变结构的分析，SEM 可以导出特定的参数，并进行整体模型的评鉴与分析。SEM 分析中的参数涉及了传统因素分析当中的因素载荷，也与多元回归的参数估计有关。因此，下一个问题就是讨论 SEM 进行参数估计的统计原理与方法。

在结构方程模型中，至少有下列几种方法可以用以估计各参数：工具性变量方法（instrumental variable method, IV）、二阶段最小平方法（two-stage least squares, TSLS）、无加权最小平方法（unweighted least squares, ULS）、最大概似法（maximum likelihood, ML）、一般化最小平方法（generalized least squares, GLS）、椭圆分布理

论法 (elliptical distribution theory, EDT) 或渐进分布自由法 (asymptotic distribution free, ADF)。这些不同方法的共同点, 是求取观察与估计共变结构之间差异的最小值, 来导出各参数的最佳估计数。

在各种方法当中, 最简单的方法是直接分别估计每一个方程式, 求得每一个方程式的解, 而不必参考模型当中其他方程式的条件, 称为有限讯息技术 (limited-information techniques), IV 法与 TSLS 法即属此类技术。相对之下, 完全讯息技术 (full-information techniques) 则是同时考虑模型当中所有方程式, 以获得各参数的估计数, 所谓完全讯息是指充分运用模型当中的所有计量信息, 以获得最理想的估计数, 一般常用的 ULS、GLS、ML、ADF 等各种方法均属此类。

从统计原理来看, 有限讯息技术是单变量 (univariate) 估计程序。也就是说, 估计程序是各方程式单独进行; 而完全讯息技术则是多变量 (multivariate) 估计程序, 每一个参数在估计时, 同时考虑其他的方程式。因此, 有限讯息技术并没有迭代估计 (iterative estimation) 的问题, 而完全讯息技术则必须进行迭代程序, 以求取最理想的终解。一般而言, 有些学者利用 IV 法或 TSLS 法估计起始值 (starting values)。这些起始值可以用于完全讯息技术法的参数估计, 并且节省估计的时间, 提高 SEM 分析的效率。

由于有限讯息技术并不需要进行迭代估计, 因此具有速度快, 识别错误的问题较少, 不受某些高阶统计假设的限制等优点, 因此统计强韧性 (robustness) 较高。相对之下, 完全讯息技术则有较佳的统计效度, 得到的结论较为合理 (Long, 1983)。

经由迭代程序, 完全讯息技术可以在一定的重复估计程序中找到最理想的解, 此一程序称为有效的收敛 (convergence)。所得到的估计结果应该是可以被理解的, 称为可接受解 (admissible solution) 或是适当解 (proper solution)。相对地, 有时候研究者所提出的模型不恰当、模型数据与观察数据相差太大, 或是起始值与最终解的差距过大, 迭代程序无法在少数几次内完成。即使完成了估计, 得到的估计解无法理解、无法解释, 称为不可接受解 (non-admissible) 或不适当解 (improper solution), 例如超过范围的标准化估计数、方差出现负值 (称为 Heywood cases)、非正定问题。由此可知, 一个具有理论与实务合理性的模型, 对于 SEM 的参数估计有着相当决定性的影响。而起始值的估计虽然是改善终解适合性的一种策略, 但是基本问题还是在待验证的假设模型之上。

一、加权最小平方策略

在结构方程模型的分析当中，参数估计的进行是透过估计协方差矩阵与观察协方差矩阵的差异极小化来完成的。而估计协方差矩阵与观察协方差矩阵的差异，可以用一个 $F(Q)$ 函数关系来表示，称为拟合函数（function for fitting covariance structures）：

$$F(Q) = (s - \sigma(\Theta))^T W^{-1} (s - \sigma(\Theta))$$

$$= \sum_{g=1}^k \sum_{h=1}^g \sum_{i=1}^k \sum_{j=1}^i w^{gh,ij} (s_{gh} - \sigma_{gh})(s_{ij} - \sigma_{ij}) \quad (3.8)$$

其中 s 是观察数据向量，也就是从样本观察到协方差矩阵（S矩阵）向量， σ 是估计协方差矩阵（ Σ 矩阵）向量，两者的差距是为残差；其中 σ 为 Θ 的函数， Θ 是指 SEM 模型中的各项参数，例如回归系数、因素载荷、方差与协方差等。 W^{-1} 用来校正参数估计的校正权数矩阵（correct weight matrix）。不同的估计策略主要差异在于所使用的 W^{-1} 矩阵不同，求取 $F(Q)$ 函数最小化的方式不同。应用此一拟合函数所进行的参数估计，统称为加权最小平方法（weighted least-squares; WLS），ULS法、GLS法、ML法都可以视为WLS法的一种特殊应用情况¹。不同的估计策略，主要差异在于所使用的 W^{-1} 矩阵不同。

由于 WLS 的应用弹性大，因此深受研究者的喜爱，但是 WLS 法有几个基本问题。第一，由于 W 矩阵是一个 $u \times u$ 的矩阵， $u = n(n+1)/2$ 。当观察值数目（ n ）增加，矩阵的规模即快速增加，造成执行 SEM 分析的操作时间与复杂度。例如当 $n=20$ ， W 矩阵共有 22155 个元素。

第二，当存在着遗漏值时，估计的进行会因为遗漏的形态而影响，因此必须使用全列排除法（listwise deletion）而非配对排除法（paired deletion），将具有遗漏值

¹ 在各种估计程序中，ULS法得到的结果不符合卡方统计的原理（其标准误的计算不同于卡方的期望值计算原理），因此ULS法已经逐渐不被纳入主要的参数估计策略。

的样本去除，此时将造成样本的流失与流失一致性的问题。

第三，配合 W 矩阵权数不同的估计法，整个 SEM 分析需要大量的样本数，动辄数百至数千人。这无疑提高了实际操作的难度。一般而言，当样本数越大，非正态对于统计估计的影响就可以缩小。但是到底多大的样本才足够呢？研究者之间并没有一定的共识。一般而言，样本数必须达到一个 SEM 模型当中自由估计的参数数目的 10 倍以上 (Bentler, 1992; Hu, Bentler & Kano, 1992)。如果样本数不足，研究者可以使用替代的估计法来估计参数。

基本上，WLS 法必须建立在一定的统计假设之上，例如当观察变量的正态分布假设违反时，统计检验正确性可能违反，WLS 的结果将被扭曲。如果将 $F(Q)$ 乘以 $N-1$ ，即可得到卡方值，适用卡方分布来决定其统计显著性。在统计程序上，卡方检验是求取细格的期望值，计算观察值与期望值的差异的平方值。因此，如果 W 矩阵的计算不符合期望理论值，那么就无法应用卡方分布来决定统计显著性。

虽然学者提出不同的方法来处理这些单变量的非正态问题，例如 Satorra 与 Bentler (1988) 发展出非正态性数据的校正技术 Satorra-Bentler Scaled χ^2 。但是，如果研究者的数据具有相当程度的偏态时，必须在进行 SEM 分析之前，先行处理测量变量不佳的现象，以及离漏值的问题。否则，仰赖校正统计，不仅耗时（例如 EQS 软件的 Robust 程序可以处理偏态问题，但是一般 SEM 耗时的 10 倍以上），有些软件甚至并未提供校正功能。校正的进行，也是在统计过程增加了不稳定的因素，着实不是 SEM 分析的有利条件。

（一）无加权最小平方法（ULS）

无加权最小平方法的原理是求取 Σ 与 S 矩阵的差异（残差矩阵）平方和的最小值，各向量元素差异值的计算并没有经过加权处理， W 矩阵当中的权数均为 1，也就是每一个向量具有相等的重要性。因此，当所有的观察变量有类似的测量尺度时（单位均相同时），适合使用此一方法。其函数为：

$$F_{\text{ULS}} = \frac{1}{2} \text{tr}[(S - \Sigma(\theta))^2] \quad (3.9)$$

以 ULS 法来进行参数估计, 即是将上述函数予以最小化估计。 $tr[(S - \Sigma(\Theta))^2]$ 表示残差矩阵当中各元素的平方和, 每一个元素具有相等的计量尺度。利用此一方法进行参数估计的优点是计算简单、快速, 但是拟合函数极小化的处理能力较差。例如 ULS 法没有考虑到每一个向量的残差因为量尺单位 (或各变量的标准差) 不同造成的异质性 (heteroscedasticity), 因此以 ULS 法估计大样本的矩阵, 其残差变异量相当大, 估计结果效果差。此外, 利用 ULS 法估计所获得的标准误, 是在正态假设成立的基础上, 如果假设违反, ULS 法的估计结果将不可信赖。

(二) 一般化最小平方法 (GLS)

一般化最小平方法的基本原理也是使用差异平方和的概念, 只是在计算每一个差异值时, 以特定的权数来加权个别的比较值。此概念最早由 Aitken (1934) 所提出, Jöreskog 与 Goldberger (1972) 随后将其应用于路径分析当中。GLS 法是由 ULS 法改进而来, 即求取下列函数的最小化:

$$F_{GLS} = \frac{1}{2} tr \{ (S - \Sigma(\Theta)) W^{-1} \}^2 \quad (3.10)$$

由 3.10 的函数可知, GLS 法在估计残差矩阵的最小平方数值时, 增加了一组权数 W^{-1} , 使矩阵中的不同向量依照权数加权后求得函数的最小化。其原始目的是在校正 ULS 法当中无法处理的残差异质性问题, 具体做法是将残差乘以观察协方差矩阵的反矩阵 S^{-1} (令 W^{-1} 为 S^{-1}), 也就是将每一个残差除以自己的协方差或方差。因此, GLS 的函数又可以写成:

$$F_{GLS} = \frac{1}{2} tr \{ (S - \Sigma(\Theta)) S^{-1} \}^2 = \frac{1}{2} tr \{ I - S^{-1} \Sigma(\Theta) \}^2 \quad (3.11)$$

二、最大概似法 (ML)

最大概似法 (Maximum Likelihood) 是 SEM 分析最常用的参数估计法。最早由 Koopmans、Rubin 与 Leipnik (1950) 提出时, 称为完全讯息最大概似法 (full-information maximum likelihood), 其后由 Jöreskog (1971) 将其纳入 SEM 分析当中。ML 法的函数式如下:

$$F_{ML} = \log|\Sigma| - \log|S| + \text{tr}(S\Sigma^{-1}) - \rho \quad (3.12)$$

上式中, ρ 为测量变量的数目 ($p+q$), Σ 为估计总体协方差矩阵。当估计矩阵与观察矩阵完全拟合时, Σ 矩阵的对数值与 S 矩阵的对数值相减为 0, 而 $\text{tr}(S\Sigma^{-1})$ 则为 $\text{tr}(I)$, 解开后数值等于 $p+q$, 亦等于 ρ , 因此后两项的和也会为 0, 使得 F_{ML} 为 0。由于 F_{ML} 是基于概率原理的非线性函数, 不容易获得参数解, 因此仰赖迭代的程序来获得参数最后终解。

最大概似法的基本假设, 是观察数据都是从总体中抽取得到的数据, 且所抽出的样本必须是所有可能样本中被选择的几率最大者。若能符合此一假设, 估计的参数即能反映总体的参数。将此一概念以函数方式表示, 则是求取估计总体协方差矩阵 Σ , 并以其反矩阵为加权参数, (即 W^{-1} 权数), 来导出拟合函数值。对于 W^{-1} 矩阵, F_{ML} 以协方差矩阵的估计倒置矩阵 Σ^{-1} 为之, 而非 GLS 法中的样本观察共变的倒矩阵。

ML 法对于概率的估计, 是基于参数必须符合多变量正态假设的假设, 将多变量正态分布的概率密度分布也一并考虑在 F_{ML} 中, 估计每一个参数从总体中加以估计的最大概率。在计算上, 各变量于正态分布下的概率密度计算矩阵式如下:

$$f(z) = (2\pi)^{-(p+q)/2} |\Sigma|^{1/2} \exp\left[-\frac{1}{2} z^{-1} \Sigma^{-1} z\right] \quad (3.13)$$

上式中, z 为观察值向量, 代表来自多变量正态分布总体的随机样本在某一变量的

测量。假设随机样本中的 N 个观察值彼此独立,联合概率密度为个别概率密度的乘积, $f(z_1, z_2, \dots, z_N) = f(z_1)f(z_2)\dots f(z_N)$, 此式又称为概似函数 $L(\Omega)$, 也就是该样本参数的概率 (likelihood) 的数学定义。将此函数取对数值便可以导出 F_{ML} 函数通式。最后求取最大的概率, 即为最佳的估计值。

ML 法有一个重要的特性即概率的估计是一个渐进的过程, 必须在样本规模较大时, 即各参数的多变量正态假设成立下, 才能得到不偏的估计数, 也就是一种渐进不偏性 (asymptotic unbiasedness), 而样本参数估计数与参数的总体之间也呈现渐进的一致性 (asymptotic consistency)。基于中央极限定理, 当样本数增加, 所有参数估计误的变小, 呈现出渐进有效性 (asymptotic efficiency)。

因此, 应用 ML 法最大的要求是变量的正态性必须有效维系。前述 ULS 法与 GLS 法也基于正态分布假设来进行估计, 因此使用这些估计程序之前, 都必须先行检验观察变量的正态化假设是否成立, 否则不宜使用这两种估计方法。Raykov 与 Marcoulides (2000) 指出除了检验观察变量的偏态与峰度, 应利用一些统计指标检验多变量正态性, 例如 Mardia 多变量峰度系数 (Mardia multivariate kurtosis coefficient)。如果当观察数据呈现明显违反正态化假设时, 同时样本数较大时, 可以采用 EDT 法或渐进分布自由法 (ADF) 进行参数估计。

三、渐进分布自由法 (ADF)

渐进分布自由法是由 Bollen (1984) 所提出的一种无须正态假设为基础的参数估计法, 由于不需考虑正态分布的问题, 因此称为分布自由 (distribution free)。ADF 法也可以视为是 WLS 法的一种特例, 利用特殊的 W^{-1} 权数, 来消除多变量正态假设的影响。其拟合函数公式如下:

$$F_{ADF} = \frac{1}{2}(\kappa+1)^{-1}tr\{[S - \Sigma(\Theta)]W^{-1}\}^2 - \delta\{[S - \Sigma(\Theta)]W^{-1}\}^2 \quad (3.14)$$

其中 κ 、 δ 为观察变量的峰度 (kurtosis)。又

$$F_{\text{ADF}} = [s - \sigma(\theta)]' W^{-1} [s - \sigma(\theta)] \quad (3.15)$$

其中 W 权数矩阵包括不同类型权数的整合:

$$w_{ijkl} = \sigma_{ijkl} - \sigma_{ij}\sigma_{kl} \quad (3.16)$$

σ_{ijkl} 为峰度, σ_{ij} 为协方差。透过上述程序, F_{ADF} 的计算考虑了非正态分布的影响, 因此 F_{ADF} 可以不受分布正态性的影响。但是, 使用 ADF 法有下列限制:

第一, F_{ADF} 函数所处理的是测量变量的峰度, 因此利用 ADF 分析必须使用原始数据, 无法利用共变或相关矩阵输入法。其次, ADF 法的计算过程繁复。在导出协方差矩阵过程中, 针对各变量的非正态性进行校正处理, 因此需要更长的计算时间, 耗费较多的计算机内存, 使用者必须配有较佳的硬件才能有效处理 ADF 的计算过程。同样, 因为参数估计涉及四级动差的运算, 因此样本规模必须一定规模才能使估计结果趋于稳定 (West, Finch, Curran, 1995), 通常 ADF 法的样本数甚至要高达 2500 人以上才趋于稳定。

上述的限制, 不仅出现在 ADF 法, 也出现在各种估计方法中。可见在技术层次, 还有诸多的因素考验着 SEM 分析的稳定性与实际的可操作性。到目前为止, SEM 研究者还不断地为改善估计的方法、提升估计的效率各自努力, 并没有哪一种策略可以说是最佳的估计策略。使用者应斟酌各自的状况, 挑选最适策略, 或是尝试使用不同的策略, 比较出最佳的结果。

第四节 参数估计的相关议题

一、参数估计与样本数的关系

为了使 SEM 分析能够在正态化假设成立的条件下进行,维持一定规模的样本数是必要的。一般而言,最大概似法的参数估计,样本数需到达 500 人,正态假设的共变结构分析才能够维系;在 500 人以下时,GLS 方法较佳 (Hu, Bentler & Kano, 1992)。

Bentler 与 Yuan (1999) 提出了一个统计量 Yuan-Bentler's T , 稍微修正了 ADF 法的计算过程,发现在样本规模为 60 到 120 之间时,仍有相当稳定的估计结果。公式如下:

$$T = \frac{[N - (p^* - q)]T_{ADF}}{[(N - 1)(p^* - q)]} \quad (3.17)$$

其中 N 为样本数, p 为测量变量的数目, $p^* = [p(p+1)]/2$ 为观察变量所提供 SEM 分析的数据数 (data points), q 为估计参数数目, T_{ADF} 为基于 ADF 方法所估计出来的统计量。

当正态性假设违反时,以 ML 法与 GLS 法来估计参数,样本规模必须达 2500 以上,参数估计才能趋于稳定 (Hu, Bentler, & Kano, 1992)。样本越少时,GLS 法较 ML 法稍微好一点。ADF 法在样本数 2500 以下时表现不理想。如果误差项独立假设违反,ML 法与 GLS 法估计效果在任何样本规模下均不理想,ADF 法在样本数大于 2500 时表现理想。EDT 法比 ML 法、GLS 法与 ADF 法要好,在中度到大型样本数的情况下,度量化 ML 法 (scaled ML, 为修正正态化不足后的最大概似法) 表现最好。

总而言之,测量变量的正态性假设与独立性假设是影响 SEM 参数估计与假设检验的重要因素。ML、scaled ML、GLS 方法适用于中到大规模样本的 SEM 分析,scaled

ML 法尤其依赖计算机的校正,因此,如果考虑成本因素,ML 与 GLS 是适合的参数估计方法,尤其是 ML 法,已经成为 SEM 分析参数估计的主要方法。如果配合 ML 的度量化校正,非正态的威胁可以被加以修正,使得参数估计更为稳定。但是由于度量化 ML 法必须针对所有的观察变量进行调整,即使使用大型计算机工作站,估计过程也非常耗费时间。因此,使用者可以先利用 ML 法进行整体 SEM 模型检验,当进行参数估计时,则使用度量化 ML 法,将可以得到最佳的效果。在小样本下,可以使用 Yuan-Bentler 所发展的 T 统计量。

二、模型参数估计的迭代

由于参数的数值大小是被估计出来的,因此,估计的程序必须能够不断去尝试各种可能的数值以获得最适合的结果(使估计矩阵与观察矩阵的差异最小化)。如果以人为的方式来计算最佳解,是一件旷日废时的工作。但是透过分析软件,计算机可以快速地反复进行估计。每一次的估计都可以稍微缩小观察矩阵与理论模式矩阵的差异,直到无法进一步有效改善模型的拟合度,计算机即自动停止参数估计,达成收敛,得出一组优化的参数值,称为终解(final solution),此一过程是所谓的迭代估计(iterative estimation)。有时 SEM 的迭代估计在一定的次数内无法获得有效收敛以获得终解,即使得到任何参数估计值也是不值得采信的。

一旦获得终解之后,SEM 估计也可以获得每一个参数的估计标准误(standard error),用以反映每一个参数估计的可能波动范围,也就是估计误差的大小。利用参数估计值除以标准误,可以计算出统计检验值(t 值),再配合 t 分布,可以进行参数的显著性检验。一般而言,如果 t 值的绝对值大于 2,则该参数即可达到 .05 显著水平。在样本数低于 30 之时,样本数越小, t 值要越大才能超越显著水平的门槛。读者可以参酌 t 分布表来获得特定自由度之时的显著水平关键值。

三、非正定问题

曾经执行过 SEM 参数估计的使用者，多曾遇到一种称之为非正定 (non-positive definition) 的技术问题，造成统计分析不正常终止的现象。所谓正常定义是指反映分析过程数学运算程序的合理性。在 SEM 分析中，实际观察到的协方差矩阵与理论导出的协方差矩阵必须是正常定义或非歧异性 (nonsingular)，否则将使 SEM 的数学计算陷入无意义的困境。

SEM 矩阵估计的非正定问题，通常发生于下列几种状况：

第一，矩阵中对角在线的数值通常为观察变量的方差或自身相关系数 (1.00)，其数值必须为正值，否则违反数学原理。

第二，对角在线的数值是其他元素的基本条件，也就是说，方差是计算协方差的基础，一定的方差的条件下，协方差有其存在的合理范围。因此，当矩阵中有超出一定合理的范围时，即违反数学原理，称为三角不均等条件 (triangular inequality condition)。

方差为负值或数值协方差超过合理范围的现象，有一个可能性是数据输入错误所造成的。SEM 分析可以直接读取协方差矩阵，不必从原始数据逐个读取。但是，可能在矩阵数据处理过程中，由于操作者的疏失产生错误，造成不合理的现象。

此外，当数据库当中存在着遗漏值，且呈现非系统性遗漏现象时，如果使用配对排除法或以其他方式填补遗漏值，都可能造成不合理数值的出现。LISREL 第八版软件提供了一种平滑程序 (smoothing procedure)，利用脊状估计法指令 (RIDGE)，可以将不正常负数调整为正值。

第三，为了使 SEM 的参数估计可以顺利得到反矩阵，因此矩阵必须符合非歧异的要件。如果在倒置过程中，分母项为零，将造成数学的无意义除式，将致使分析中止。通常此一问题会发生在变量之间具有高度相关的情况下，也就是变量间线性依赖或共线性 (collinearity) 的问题。当某一个变量是其他变量的线性整合时，反矩阵即可能出现非正定问题。

例如，当观察变量中，有某一个变量是其他变量的组合变量（其他变量的加减乘除），此一变量便会与参与组合的变量形成高相关，造成非正定问题。或者是当样本数少于观察变量数，或观察变量存在着特殊的偏离值或极端值，也可能造成变量间的线性依赖问题。这些问题与研究者的测量程序与数据准备程序是否严谨有关。这再次凸显了测量工作的优先性，与事前妥善的过滤与检查数据的重要性。

整体而言，造成 SEM 参数估计的非正定问题的原因不仅是上述三类原因，而可能是其他问题（例如不恰当的假设模型、不恰当的起始值、模型的识别性问题等等），或是多重原因所造成的影响（Schumacker & Lomax, 1996）。研究者除了提高测量的质量，落实数据检查与数据准备，还必须详加研读 SEM 各项操作程序的要领与基本原理，避免错误地使用。如果非正定问题仍然没有办法解决，最好的方式是重新建立模型，重新进行 SEM 分析。此外，就必须依赖研究者操作 SEM 的经验与功力了。

例 1 某工厂生产某种产品，其成本函数为 $C(x) = 0.01x^2 + 0.5x + 100$ ，其中 x 为产量， $C(x)$ 为成本。求该工厂生产 100 单位产品时的平均成本。

解：平均成本函数为 $\bar{C}(x) = \frac{C(x)}{x}$ 。将 $C(x)$ 代入，得 $\bar{C}(x) = \frac{0.01x^2 + 0.5x + 100}{x}$ 。将 $x = 100$ 代入，得 $\bar{C}(100) = \frac{0.01 \times 100^2 + 0.5 \times 100 + 100}{100} = \frac{100 + 50 + 100}{100} = \frac{250}{100} = 2.5$ 。因此，该工厂生产 100 单位产品时的平均成本为 2.5。

例 2 某工厂生产某种产品，其成本函数为 $C(x) = 0.01x^2 + 0.5x + 100$ ，其中 x 为产量， $C(x)$ 为成本。求该工厂生产 100 单位产品时的边际成本。

解：边际成本函数为 $C'(x)$ 。对 $C(x)$ 求导，得 $C'(x) = 0.02x + 0.5$ 。将 $x = 100$ 代入，得 $C'(100) = 0.02 \times 100 + 0.5 = 2 + 0.5 = 2.5$ 。因此，该工厂生产 100 单位产品时的边际成本为 2.5。

第四章

模型拟合评鉴

在完成了参数估计之后，结构方程模型分析的另一个重要工作是进行模型的评估与检验，以决定研究者所提出的假设模型是否能够用以描述实际观察到的变量关系，此一过程称为模型拟合评鉴（model-fit evaluation）。由于 SEM 模型包括多种参数估计，整体模型的考验涉及繁琐复杂的计量程序，用以反映模型拟合程度的方式也很多，读者应熟知这些拟合度评估的原理与差异，才能获得正确合宜的评估结果。

基本上，结构方程模型的拟合度评估与其他多变量统计（例如对数线性模型分析、逻辑回归分析）的做法类似，都是以不显著的卡方值来反映理想的模型拟合度。此做法的基本概念是从抽样理论出发，因为 SEM 模型的分析是基于某一个特定样本的观测数据，以特定的假设模型为架构，推导估计出所有的参数数据。所以，将 SEM 模型的分析结果与样本的抽样分布进行比较，并利用卡方分布检验模型是否与观察数据拟合，是一个最早被接受的做法。

虽然以卡方分布进行模型拟合度检验符合抽样理论的基本原理，但是以卡方检验进行统计决策却有诸多限制，直接以卡方检验决定 SEM 模型的适合与否备受质疑。例如随着样本数的增加，卡方值也会自然增加，无关乎模型本身的优劣；但是在样本很小时，模型估计得到的统计量分布则不一定是卡方分布，若以卡方分布进行显著性检验，又会产生偏误的结果（Bentler, 1988）。基于这些问题，学者们于是开始研发各种替代性的模型拟合指数，到目前为止，各种不同的指数仍不断被发表出来。已被发表出来的指数是否合理、恰当也不断有专文在质疑与争辩。本章仅就一些目前常用的指数做一介绍。对于它们的学术评价则有赖读者自行查阅当代学术界最新的看法。

第一节 模型评鉴的基本概念

承如先前所提及，结构方程模型是一套用以检验特定假设模型的统计方法学。因此，SEM 最主要的一个目的是在验证研究者所提出的理论或概念架构是否具有实证的意义。整个 SEM 的分析程序，都离不开研究者所提出的假设模型，因而研究者是否可以在提出研究问题的第一时间，就能够透过理论推导与文献检阅过程，选择

适当的研究变量，提出有意义的研究假设去说明变量的关系，进而发展出适切理想的假设模型，即成为 SEM 模式的计量检证程序是否可以顺利完成的一个基本条件。换句话说，SEM 分析技术只是一套统计的方法与分析的策略，SEM 的本身并无法创造理论或知识，而是需要研究者以其智慧去整理前人所发展出来的理论或知识，建构出一套适当的概念模型，然后再以 SEM 技术协助研究者完成模型的分析与讨论。

一、测量质量与模型评鉴

要能够顺利完成 SEM 的计量分析，除了假设模型的发展，还有赖于严谨的抽样与测量程序，以获得稳定、有效的观察数据来评估 SEM 模式的適切性。

在计量领域中，测量与分析可以说是两套独立的程序，如果抽样有所偏颇或测量的过程带有大量误差，不仅在检验过程当中会产生统计上的诸多问题，更可能导致错误的结论。换句话说，当观察变量的信效度不佳、测量质量低下时，SEM 的分析也不会有理想的结果，而且此一不理想的结果，可能完全与假设模型的优劣无关，但是却可能让人们误将研究结果的问题指向假设模型的本身。因此，在进行 SEM 分析之前，研究者必须具体指出变量的操作型定义，谨慎小心地处理测量的信效度问题，以避免上述困境的发生。

二、模型评鉴的假设检验

SEM 模型评鉴的一个重要概念，是 SEM 分析只能用来评估研究者所提出的假设模型是否適切。但是，究竟何者才是真正能够反映真实世界的变量之间关系的模型，这一个结论并不能够从模型评鉴的程序中得出。因为除了研究者所提出的模型之外，同样的一组观察变量可能有许多不同的模型组合。这些基于同样观察数据基础的假设模型可能都有理想的拟合度，SEM 分析并无法区辨这些计量特征类似的理论模式何者为真。SEM 的使用者不但必须谨记统计学本身的限制，也必须避免陷入过度推论的陷阱当中。

SEM 分析策略与其他推论统计方法的一个明显差异,是以支持虚无假设为模型拟合度存在的证据。一般来说,推论统计的各种统计检验多以推翻某一寻常假设(虚无假设),以获得研究者所提出的特定假设(对立假设)为真的结论为目标。当虚无假设不成立之时,研究者即可以据以“证明”所提出的假设是有其存在的意义与价值。

但是,在 SEM 中,假设模型的適切性是以虚无假设的形式存在,虚无假设代表 SEM 假设模式与实际观察所得数据相符合,对立假设则表示理论模式不能反映观察数据。很明显的,基于此一假设检验逻辑所获得的虚无假设成立的统计决策(也就是我们可以主张被检验的 SEM 模型是一个适当的模型时),只能说明该 SEM 假设模型与观察数据没有不同,但是却无法“证明”该 SEM 假设模型是否有别于任何其他的模型。这也就是为何 SEM 的研究者无不挖空心思去发展计量的技术,以区辨、评估 SEM 模型的好坏优劣。

三、参数估计与模型评鉴

从 SEM 的执行技术来看,假设模型的检验是在完成参数估计之后才进行的工作。因此,评估模型適切性的第一步,应是 SEM 的模型与各项参数能够顺利地被识别、收敛、估计。然后藉由模型参数的质量与适当性的检验,对于假设模型有一个初步的筛检。

具体来说,研究者可以逐一检视参数估计的结果,检查每一个参数的正负号、数值大小是否符合理论预期,或是检查测量误差的大小,分析这些残差项当中是否透露某些变量的测量质量不佳的讯息。如果某些变量的测量误差过于严重,研究者应先行解决测量的问题,重新检讨参数的估计,而非进入模型评鉴的程序。

第二节 模型评鉴的方法

一旦 SEM 假设模型中的每一个参数被顺利估计之后，SEM 即可以进行整体模型的评估。透过不同的统计程序或拟合指数（goodness-of-fit index）的计算，研究者可以分析假设模型与实际观察数据的拟合情形。如果模型拟合度不理想，代表研究者所提出的假设模型可能存在某些问题，可能是模型的设定、或是参数的估计、或是技术上的问题导致假设模型无法与观察数据拟合。此时研究者可以应用模型修饰的原则，调整假设模型的参数估计内容，重新加以估计，直到模型拟合度达到理想的水平。在此过程必须反复检验模型拟合指数，可见模型拟合度评估在 SEM 分析当中的重要性。

以 LISREL 为例，执行完毕后，从报表中可以得到一个模型拟合度检验报告，其中罗列了各式各样常用或特殊的模型评鉴指数。表 4.1 即是根据本书范例一（验证性因素分析）所得到的模型评鉴指数数据。以下我们将逐一说明各数据的意义与使用方法。

表 4.1 验证性因素分析（范例一）的模型拟合度分析数据

1	Degrees of Freedom = 120
2	Minimum Fit Function Chi-Square = 247.75 (P = 0.00)
3	Normal Theory Weighted Least Squares Chi-Square = 229.25 (P = 0.00)
4	Estimated Non-centrality Parameter (NCP) = 109.25
5	90 Percent Confidence Interval for NCP = (70.31 ; 156.01)
6	Minimum Fit Function Value = 0.71
7	Population Discrepancy Function Value (F0) = 0.31
8	90 Percent Confidence Interval for F0 = (0.20 ; 0.45)
9	Root Mean Square Error of Approximation (RMSEA) = 0.051
10	90 Percent Confidence Interval for RMSEA = (0.041 ; 0.061)
11	P-Value for Test of Close Fit (RMSEA < 0.05) = 0.42
12	Expected Cross-Validation Index (ECVI) = 0.95
13	90 Percent Confidence Interval for ECVI = (0.84 ; 1.08)
14	ECVI for Saturated Model = 0.98
15	ECVI for Independence Model = 21.79
16	Chi-Square for Independence Model with 153 Degrees of Freedom = 7569.35
17	Independence AIC = 7605.35
18	Model AIC = 331.25
19	Saturated AIC = 342.00

20	Independence CAIC = 7692.80
21	Model CAIC = 579.01
22	Saturated CAIC = 1172.71
23	Normed Fit Index (NFI) = 0.97
24	Non-Normed Fit Index (NNFI) = 0.98
25	Parsimony Normed Fit Index (PNFI) = 0.76
26	Comparative Fit Index (CFI) = 0.98
27	Incremental Fit Index (IFI) = 0.98
28	Relative Fit Index (RFI) = 0.96
29	Critical N (CN) = 223.91
30	Root Mean Square Residual (RMR) = 0.054
31	Standardized RMR = 0.049
32	Goodness of Fit Index (GFI) = 0.93
33	Adjusted Goodness of Fit Index (AGFI) = 0.90
34	Parsimony Goodness of Fit Index (PGFI) = 0.65

一、卡方检验

(一) 卡方检验的原理

SEM分析当中，最常用的模型评鉴方式是卡方检验 (χ^2 test)。在SEM中，卡方值是由拟合函数所转换而来的统计量，反映了SEM假设模型的导出矩阵与观察矩阵的差异程度¹。卡方值的导出式如下：

$$T = (N - 1)F_{\min} \tag{4.1}$$

T 代表模型拟合度的检验值，其性质与卡方值相同，因此可以视为卡方值； N 为样本数， F_{\min} 表示以各种不同参数估计方法（例如ML，GLS，ADF等）所得到的拟合函数最小函数估计值，该数值也反映了优化的参数统计量（假设模型与观察数据差异的最小值）。在符合卡方分布的条件下，可以对于 T 值进行卡方检验来检验其显著性。此时，虚无假设为理论矩阵与观察矩阵没有差异 ($H_0: \Sigma = \Sigma(\theta)$)（模型拟合度

¹ LISREL的结果报告总共可以产生C1至C4四种卡方值，C1是导自最小契合函数，C2导自正态化加权最小平方契合函数，C3是Satorra-Bentler量尺化卡方统计量 (SB scaled chi-square statistic)，C4是导自渐近共变矩阵。计算原理请参考LISREL使用手册。

良好)。当 T 值到达显著水平,代表虚无假设不成立,模型拟合度不佳;反之,当 T 值未达显著水平,代表虚无假设成立,模型拟合度良好。

以表 4.1 的数据为例,拟合函数转换为卡方值为 247.75 ($p=.00$) (第 2 行),自由度为 120,表示模型的拟合情形不佳。由于范例一的样本数为 350 人,根据计算式: $247.75=(350-1) \times .71$,可以得出以最大概似法所得到的拟合函数估计值 F_{\min} 为.71 (第 6 行)。

当使用 GLS、ML、DWLS 等估计程序时, LISREL 报表另外报告了一个基于正态假设的加权最小平方卡方值 (Normal Theory Weighted Least Square Chi-square; WLS χ^2), 其数值为 229.25。该值具有统计显著性,表示模型的拟合度不理想,也就是由假设模型所导出的协方差矩阵 $\Sigma(\theta)$ 与观察矩阵 Σ 相等的虚无假设被推翻。

(二) 卡方自由度比

在卡方检验的概念下,我们可以理解自由度越大的模型在卡方统计检验上越处于不利的地位。因此,如果有两个模型同时进行 SEM 分析,得到不显著的卡方值之时,自由度越大的模型越有能力去反映真实的数据,这就是 SEM 中常见的简效原理的统计原理。由此可见,如果利用卡方数据去检验模型拟合度之时,除了依循卡方值越小越好的统计显著性原理之外,还需考虑自由度大小的影响。

在 SEM 分析当中,可以计算出一个卡方自由度比 (χ^2/df),称为正规卡方值 (normed chi-square),来进行模型间拟合度的比较。卡方自由度比越小,表示模型拟合度越高,反之则表示模型拟合度越差。一般而言,卡方自由度比小于 2 时,表示模型具有理想的拟合度 (Carmines & McIver, 1981)。

以表 4.1 的数据为例,卡方自由度比为 $247.75/120=2.06$,若以正态化 WLS 卡方来计算,则为 $229.25/120=1.91$,已达 2.00 的门槛,显示如果考虑了模型的复杂度(自由度越大,模型越复杂),该范例的模型拟合度尚称理想。由于 SEM 分析深受非正态数据的影响,因此在进行 LISREL 分析时,多以 WLS 卡方值来描述模型的拟合度¹,建议

¹例如 LISREL 在估计非中央性参数 (NCP) 时,即使用 WLS 卡方值,而非最小契合函数卡方值。

研究者在报告最终结果时，亦使用WLS卡方值。

（三）卡方检验的相关议题

以卡方检验评估理论模式的適切性应注意几个问题。第一，承如先前所提到，SEM当中，被检证的假设模型若能够有效反映实际观察到的数据，其虚无假设应成立，而非对立假设。因此，利用卡方分布进行假设模型的统计检验，并无法推翻不良的模型以支持特定的模型，只能确认虚无假设是否成立。此程序无法检验假设模型的特定意义，只能说明假设模型并非不能反映观察资料，但是不能证明假设模型是否真的反映了它所代表的理论概念。

进一步的，由于卡方分布受到自由度的影响，当自由度越大，卡方值越大。也就是说，当自由度越大时，所欲估计的参数数目越多，影响一个假设模型的因素就越多，造成假设模型拟合度不佳的可能性越大，此时卡方值越不能够用来表明理论模式是否能够表明观察数据的程度。

此外，卡方分布也与样本数有关。当样本越大，所累积的卡方值也就越大。大样本虽然提高了观察数据的稳定性，却也造成卡方值扩大的效果。虚无假设被拒绝的概率与自由度及样本数具有正比函数关系。因此，当利用卡方分布检验SEM模型时，会因为参数数目与样本数的技术特性影响假设模型的拟合度检验。这也是一般SEM使用者舍卡方就其他拟合指数的主要原因。

另一方面，基于中央极限定理，当样本够大时，从样本上获得的参数估计统计量应会越趋接近中央极限卡方分布，称为中央卡方分布（central chi-square distribution）。中央卡方分布代表假设模型参数是否能够拟合观测数据的卡方真分数。由SEM分析当中所获得的 T 值除了可以进行卡方检验来检验其显著性，亦可以与中央卡方分布进行差异比较以决定 T 值的统计意义。

二、模型拟合指数

除了卡方值与卡方检验之外，SEM使用不同的指数检验模型的拟合度。这些指

数可能基于卡方统计量,只是修正了卡方统计的某些限制,或以不同的替代性模型作为参照,使模型拟合度的程度能够被真实反映出来。最常见的模型拟合指数有 GFI、AGFI、NFI、NNFI 等几种,以下分别介绍之。

(一) GFI 与 AGFI 指数

GFI 指数即是拟合指数 (goodness-of-fit index) 的缩写,类似于回归分析当中的可解释变异量 (R^2) (Tanaka & Huba, 1989),表示假设模型可以解释观察数据的方差与协方差的比例。

$$GFI = \frac{tr(\hat{\sigma}^2 W \hat{\sigma}^2)}{tr(s' W s)} \quad (4.2)$$

上式中,分子是理论假设模型的协方差所导出加权方差和,分母是样本实际观察所得到的协方差导出的加权方差和, W 是加权矩阵。由于模型导出值会小于实际观测值,因此 GFI 是小于 1 的比值。GFI 值越接近 1,分子分母越接近,表示模型拟合度越高。相对地,当 GFI 数值越小,分子分母差距越大,表示模型拟合度越低。

AGFI (adjusted GFI) 则类似于回归分析当中的调整后可解释变异量 (adjusted R^2)。在计算 GFI 系数时,将自由度纳入考虑之后所计算出来的模型拟合指数。当参数越多时,AGFI 指数数值越大,越有利于得到理想的拟合度结论。公式如下:

$$AGFI = 1 - \frac{1 - GFI}{1 - \frac{t}{DP}} \quad (4.3)$$

公式 4.3 中的 t 为估计参数数目, DP 为观察资料数。 GFI 与 $AGFI$ 均具有标准化的特性,数值介于 0 至 1 之间,数值越大(越接近 1)表示拟合越佳,越接近 0,表示拟合越差。一个能够拟合观察数据的 SEM 模型,其 GFI 与 $AGFI$ 都会非常接近 1.00,

一般需要大于.90 才可以视为具有理想的拟合度 (Hu & Bentler, 1999)。以表 4.1 为例, $GFI=.93$, $AGFI=.90$, 皆达.90 门槛, 表示模型的拟合情形理想。由于范例一的 $DP=171$, 自由估计参数=51, 因此可以计算出 $AGFI$ 数据:

$$AGFI = 1 - \frac{1 - .93}{1 - \frac{51}{171}} = .90$$

到目前为止, 并无这两个指数的统计概率分布, 因此无法直接针对 GFI 与 $AGFI$ 进行显著性检验。

(二) PGFI 指数

GFI 指数的另一种变形是 $PGFI$ 指数 (Parsimony Goodness-of-fit Index; James, Mulaik, & Brett, 1982), 其计算式如下:

$$PGFI = \left[1 - \left(\frac{t}{DP} \right) \right] GFI \quad (4.4)$$

从计算式可以看出, $PGFI$ 指数考虑到了模型当中估计参数的多寡, 可以用来反映 SEM 假设模型的简效程度 (degree of parsimony)。其判断原理与 GFI 指数相似, $PGFI$ 指数越接近 1, 显示模型越简单。但是, 根据其计算式, 可以发现除非估计参数的数目远小于观察资料数, 否则 $PGFI$ 指数会远小于 GFI 指数数值。

Mulaik 等人 (1989) 指出, 一个良好的模型, $PGFI$ 指数大约在.5 以上都是可能的, 以表 4.1 为例, 范例一的 $PGFI$ 指数为.65, 表示拟合度佳。其计算式如下:

$$PGFI = \left[1 - \left(\frac{51}{171} \right) \right] .93 = .65$$

(三) NFI 与 NNFI 指数

另外两种相当常用的拟合指数是正规拟合指数 (normed fit index; NFI) 与非正规拟合指数 (non-normed fit index; NNFI) (Bentler & Bonnet, 1980), 这两种指数是利用嵌套模型比较原理所计算出来的一种相对性指数, 反映了假设模型与一个观察变量间没有任何共变假设的独立模型的差异程度。

SEM模型中, 将观察变量之间设定为没有任何共变情况所得到的独立模型 (independent model), 可以说是利用同一组观察变量可能组合而成的无数个假设模型当中最基本的一种状况。在概念上, 可以作为所有其他模型的基准模型 (baseline model)。独立模型表示了拟合状况最不理想的一种模型, 反映了所有的观察变量之间没有任何关联, 且自由度最大, 故以独立模型导出的卡方值 (χ^2_{indep}) 是所有可能模型的卡方值之最大值。而其他所有的模型 (称为比较模型, 以 χ^2_{test} 表示), 都是从虚无模型加以延伸的嵌套模型, 卡方值会较独立模型的卡方值为低, 可以与虚无模型相比较:

$$NFI = \frac{\chi^2_{indep} - \chi^2_{test}}{\chi^2_{indep}} \quad (4.5)$$

NFI 指数的原理是计算假设模型的卡方值 (χ^2_{test}) 与虚无模型的卡方值 (χ^2_{indep}) 的差异量, 可以视为是某一个假设模型比起最糟糕模型的改善情形。以表 4.1 的数据为例, 虚无模型的卡方值为 7569.35 (独立模型的卡方值, 位于第 16 行, 自由度为 153), NFI 为 .97 (第 23 行) 的数据计算过程如下:

$$NFI = \frac{7569.35 - 247.75}{7569.35} = .97$$

根据研究发现, 在小样本与大自由度时, 对于一个拟合度理想的假设模型, 以

NFI 指数检验拟合度会有低估的现象 (Bearden, Sharma, & Teel, 1982)。所以, 其他研究者提出了另一个 NNFI 指数, 考虑了自由度的影响, 类似于前述 AGFI 指数对 GFI 指数的调整, 由此可以避免模型复杂度的影响, 其计算式如下:

$$NNFI = \frac{\chi_{indep}^2 - \frac{df_{indep}}{df_{test}} \chi_{test}^2}{\chi_{indep}^2 - df_{indep}} \quad (4.6)$$

调整后的 NNFI 指数改善了 NFI 的问题, 但是却使得 NNFI 有时会有超越 0 至 1 范围的数值出现, 显示 NNFI 的波动性较大。同时, NNFI 系数值可能会较其他指数来得低, 使得当其他指数显示模型是拟合的情况下, 其却显示模型拟合度不理想的矛盾结论 (Anderson & Gerbing, 1984)。表 4.1 中 NNFI 指数的计算过程如下:

$$NNFI = \frac{7569.35 - \frac{153}{120} \times 247.75}{7569.35 - 153} = .98$$

(四) IFI 指数

Bollen (1989) 提出了一个增量拟合指数 (incremental fit index; IFI) 来处理 NNFI 波动的问题以及样本大小对于 NFI 指数的影响, 公式如下:

$$IFI = \frac{\chi_{indep}^2 - \chi_{test}^2}{\chi_{indep}^2 - df_{test}} \quad (4.7)$$

一般正常情况下, IFI、NFI 与 NNFI 值都会介于 0 至 1 之间, 数值越大表示拟合度越佳, 同时系数值需大于 .90 才可以视为具有理想的拟合度 (Hu & Bentler, 1999)。由表 4.1 的数据来看, 范例一的模型拟合度理想。

$$IFI = \frac{7569.35 - 247.75}{7569.35 - 120} = .98$$

三、替代指数

替代指数 (alternative index) 与模型拟合指数的主要不同在于替代指数不是以卡方统计量的假设检验进行模型拟合度的评估。替代性指数的基本想法, 是假设模型与实际观察矩阵的比较不是全有全无的概念。也就是说, 模型拟合度的检验并非针对假设模型导出矩阵与观察矩阵是否相同的这一个虚无假设进行检验。因为观察数据本身是否能够反映真实变量的关系也是有待考验的问题, 若以观察数据矩阵作为比较基准并不一定能够反映假设模型的优劣与否。例如, 观察数据本身的测量质量不良 (信效度不佳), 就可能影响观察矩阵的内容。因此, 替代性指数不再关心虚无假设是否成立, 而是去直接估计假设模型与由抽样理论导出的卡方值的差异程度。

也即由于替代指数是直接估计被检验模型与理论分布的差异程度, 因此可以在中央极限定理的基础上, 以区间估计的概念进行显著性检验。其做法是考虑抽样误差对于指数估计的影响, 将所计算出来的指数转换成特定置信水平下 (如 90%) 真实指数出现的范围。优点是可以反映抽样误差的影响。

(一) 非集中性参数 (NCP)

主要的一种替代指数是非集中性参数 (non-centrality parameter; NCP)。Steiger (1990) 称NCP为模型不拟合的自然量数 (natural measure of badness-of-fit of a covariance structure model), 其原理类似于离散量数的测量, 即计算SEM模型估计得到的卡方统计量, 距离理论预期的中央卡方分布的离散程度 (discrepancy between Σ and $\Sigma(\theta)$)。此一作为中央点的卡方分布表示最佳的模式, 当模型越不理想, 距离此一中央分布越远, 以几个标准差 (δ) 来表示 (即NCP数值)。越大的NCP值代表模

型越不理想,当NCP为0时,代表模型具有完美的拟合度¹。在LISREL报表中,除了NCP值之外,也报告NCP在90%置信水平下的估计范围。当此一区间涵盖了0,亦表示模型拟合度具有完美的拟合度。以表4.1的数据为例,NCP数值为109.25,90%置信区间为70.31至156.01,并未涵盖0。因此,以NCP指数来判断,显示模型的估计数距离中央卡方分布有一段距离。

(二) RMSEA 指数

另一个重要的替代性指数为平均概似平方误根系数 (root mean square error of approximation; RMSEA) (Browne & Cudeck, 1993), 其计算式如下:

$$\text{estimated RMSEA} = \sqrt{\frac{\hat{F}_0}{df_{test}}} \quad (4.8)$$

其中 \hat{F}_0 是被检验模型的卡方值减去自由度再除以样本数的函数值:

$$\hat{F}_0 = \frac{\chi^2_{test} - df_{test}}{N} \quad (4.9)$$

由上式可知, RMSEA 系数不受样本数大小与模型复杂度的影响, 当模型趋近完美拟合时, \hat{F}_0 接近 0, RMSEA 指数亦接近 0。RMSEA 指数越小, 表示模型拟合度佳。

由计算原理可以看出, RMSEA 与 CFI 及 NFI 不同的是 RMSEA 指数在比较理论模型与完美拟合的饱和模型的差距程度。数值越大代表模型越不理想, 数值越小代表模型拟合度越理想。Hu 与 Bentler (1999) 建议 RMSEA 指数低于.06 可以视为是

¹有关中央卡方分布与非中央卡方分布的统计原理与相关议题, 请参考MacCallum与Schumacker (1996) 的专著。

一个好模型，指数大于.10 表示模型不理想 (Browne & Cudeck, 1993)。McDonald 与 Ho (2002) 则建议以.05 为良好拟合的门槛，以.08 为可接受的模型拟合门槛。

以表 4.1 的数据为例， $RMSEA=.051$ ，相当接近.05 的门槛，但低于 Hu 与 Bentler (1999) 的.06 门槛，显示模型具有相当的理想性。以表 4.1 的数据为例，RMSEA 指数的获得程序如下：

$$RMSEA = \sqrt{\frac{\frac{247.75 - 120}{350}}{120}} = .051$$

近年来，RMSEA 指数相当受到重视，由于其他模型拟合指数多受到样本大小与观察变量分布（如平均数）的影响，而 RMSEA 则可摆脱这项困扰，且具有 NCP 指数反映假设模型与中央卡方分布的离散性特质，因此 RMSEA 指数近年来普遍为大家所接受 (Marsh, 1996)。但是最近的研究则指出 RMSEA 指数在小样本时有高估的现象，使拟合模型被视为不理想模型 (Bentler & Yuan, 1999)，因此在小样本时应谨慎使用 RMSEA 数值。

(三) CFI 指数

另一个常用的指数为 CFI 指数 (comparative-fit index) (Bentler, 1992)。CFI 指数反映了假设模型与无任何共变关系的独立模型差异程度的量数，也考虑到被检验模型与中央卡方分布的离散性。其计算原理是基于非中央性改善比 (the ratio of improvement in noncentrality; 假设模型距离中央卡方分布距离的移动情形)，得出一个非中央性参数 (noncentrality parameter) τ_i ， τ_i 越大，代表拟合度越不理想， $\tau_i=0$ 时假设模型具有完美适切性。其概念式如下：

$$\tau_{\text{indep.test}} = \chi^2_{\text{indep.test}} - df_{\text{indep.test}} \quad (4.10)$$

$$\tau_{\text{est.test}} = \chi^2_{\text{est.test}} - df_{\text{est.test}} \quad (4.11)$$

$\tau_{\text{est.test}}$ 为理论假设模型非中央性参数估计数, $\tau_{\text{indep.test}}$ 为虚无模型相对于假设模型的非中央性参数。根据上式, 得到 CFI 指数公式如下:

$$CFI = 1 - \frac{\tau_{\text{est.test}}}{\tau_{\text{indep.indep}}} \quad (4.12)$$

以表 4.1 的数据为例, CFI 为.98, 计算程序如下:

$$CFI = 1 - \frac{247.75 - 120}{7569.35 - 153} = .98$$

由于独立模型(观察变量之间没有任何的共变)是最不理想的模型, 任何模型一定较独立模型的拟合度为优; 因此, CFI 指数的数值也是越接近 1 越理想, 表示能够有效改善非中央性的程度, 其性质与 NFI 接近, 一般是以.95 为通用的门槛, 同时在小样本的 SEM 分析中, 用来评估模型拟合度十分稳定 (Bentler, 1995)。

(四) ECVI 与 AIC 指数

Browne 与 Cudeck (1993) 提出了一个期望交叉效度指数 (expected cross-validation index; ECVI), 扩大了非中央性参数的应用。ECVI 指数反映了在相同的总体之下, 不同样本所重复获得同一个假设模型的拟合度的期望值, 是用来诊断模型的复核效化 (cross-validation) 的良好指数。ECVI 值越小表示模型拟合度的波动性越小, 该假设模型越好。反之, ECVI 值越大表示模型拟合度在不同样本上的波动性越大, 该假设模型的拟合度越不理想¹。

在进行不同样本间的比较时, ECVI 较 RMSEA 为优之处在于 RMSEA 指数的计算

¹有关 ECVI 与 AIC 在复核效化的计算原理与应用, 请参考本书第九章的范例与说明。

不考虑样本数与自由度的影响，因此无法应用于不同样本数目的假设模型拟合度的比较。尤其当样本数很小但是自由度很大（模型很复杂时）时，ECVI 指数能够反映假设模型的適切性。当我们必须进行不同竞争模型（competing models）的比较时，或是从多个模型当中挑选一个最佳模型时，ECVI 指数与另外一个类似的指数：Akaike 讯息指数（Akaike information criterion, AIC）与 Akaike 一致讯息指数（Consistent Akaike information criterion, CAIC）（Akaike, 1987）是非常理想的指数，因为 ECVI 与 AIC 指数可以同时考虑样本数与模型复杂度两项因素的非中央性参数。

一般而言，ECVI、AIC 与 CAIC 指数越小表示模型越简效（parsimonious）。此时如果模型可以具有一定的拟合度，则越简效的模型更为理想。换句话说，这些指数可以作为模型选择的依据。可惜的是这两种指数的数值不是 0 至 1 的范围，因此仅适合进行模型间的比较，从 AIC、CAIC 或是 ECVI 的数值的绝对值大小，很难判定个别模型的理想性。

（五）CN 指数

在 LISREL 报表中，提供了一个特别的模型拟合统计量：关键样本指数（Critical N, CN）。CN 指数是由 Hoelter（1983）所提出，用以说明样本规模的適切性。其原理是估计若要产生一个适当的模型拟合度（不显著的卡方统计量），所需要的样本数为多少（Hu & Bentler, 1995）。Hoelter（1983）认为当 CN 指数大于 200 时，表示该模型可以适当地反映样本的数据。Byrne（1998）主张除了 200 是一个门槛之外，一个研究的样本数需大于 CN 指数所估计出来的样本数，该 SEM 分析才具有合理性。

四、残差分析指数

除了整体模型拟合度的检验，个别参数的好坏与否的检测与说明也是 SEM 当中一个重要的工作。例如在一个 SEM 模型当中，可能有某一个测量模型的结构非常不理想，观察变量的测量误差非常大，这一个部分的估计可能导致整个模型的不良拟合，此时可以透过残差分析检视 SEM 模型特定参数的设定是否理想。

一般而言, SEM分析均提供两种残差的数据: 非标准化残差(un-standardized residuals)与标准化残差(standardized residuals)。非标准化残差就是假设模型与观察数据之间差距的原始量数, 也就是参数估计无法反映实际观察数据的变异量(unexplained variance or covariance)。在LISREL中, 残差原量尺估计数列于Fitted Residuals报表中。标准化残差将残差量转换为标准 z 分数, 因此残差值将大约落于 ± 3.5 的区间中¹。

使用非标准化残差了解各估计数的良好与否的优点是可以直接应用测量的原始量尺来了解残差数值大小的具体意义, 然而, 却不利于相互的比较。当不同的观察变量的量尺或单位趋近一致时, 非标准化残差的直接比较较无异议, 但是当观察变量的性质差异较大或使用不同的量尺或单位时, 残差的数值就无法直接加以比较, 而需先行转换成为标准化分数(例如 z 分数), 也就是取标准化的残差。将所有的残差置于相同的量尺单位之上, 才能显示出残差的比较意义。

SEM的标准化残差分析, 与其他统计分析(例如多元回归分析)做法类似。当标准化残差大于+3之时, 代表该估计方差或协方差不足; 当标准化残差小于-3之时, 代表该估计方差或协方差对于两个观察变量的共变有过度解释的现象。两者都会造成模型不良的拟合状况。

残差的大小反映了不良拟合的问题, 除了从数值大小来判断, 还可以通过绘图的方式观察残差变动的趋势。例如使用标准化残差分布Q图, 可以看出在每一个估计残差与期望残差的差距分布图。如果是一个好的模型, 标准化残差Q图内的观察点应呈现水平直线, 当观察点分布在水平直线之外时, 代表不寻常的估计数的存在(Jöreskog & Sörbom, 1993)。严重的残差除了可能因为该估计数的理论假设不适当, 也可能来自于非正态化的观察数据, 因此在诊断残差的大小之时, 应配合观察数据的正态性的检验, 才能得知是否是假设模型界定的问题或是观察数据本质的问题。

目前各主要SEM分析软件均提供残差均方根指数(root mean square residual, RMR)与标准化残差均方根指数(standardized root mean square residual, SRMR)来

¹ LISREL所提供的残差分析除了原始估计量与标准化残差数据之外, 还包括摘要统计量、各种散点图与茎叶图。

反映理论假设模型的整体残差：

$$RMR = \sqrt{2 \sum_{i=1}^q \sum_{j=1}^l \frac{(s_{ij} - \hat{\sigma}_{ij})^2}{q(q+1)}} \quad (4.13)$$

$s_{ij} - \hat{\sigma}_{ij}$ 代表样本（观察）与估计（理论假设模型）的方差或协方差差异。 RMR 与 $SRMR$ 越小代表模型越能拟合观察值。由于 RMR 是基于未标准化残差值所计算得出，其数值没有标准化的特性，较难解释，因此，学者多采用标准化后的 $SRMR$ 指数来评估模型的优劣。 $SRMR$ 的数值介于 0 至 1 之间，当数值低于 .08 时，表示模型拟合度佳（Hu & Bentler, 1999）。以表 4.1 的数据为例，该模型的 RMR 为 .054， $SRMR$ 为 .049，落于 .08 的门槛下，表示残差量低，模型拟合度佳。

五、拟合指数的比较与运用

（一）模型评鉴指数的比较

虽然结构方程模型的文献提供了多种不同的模型评鉴指数，然而不同的指数得到的结果往往趋近一致；因此，选择何种指数并不是一个重要的问题。但是，如果不同的指数有不一致的结果，宜考虑假设模型的适当性，重新加以分析。

一般来说，使用者可以根据自己的偏好与分析软件所提供的指数类型评估理论假设模型的拟合度。有时某些期刊审查人对于 SEM 的使用方法以及指数的运用有独特的立场，而要求报告特定的指数与系数。一般而言，在 SEM 技术领域对于指数的优劣与选择方法并无一致的共识。目前最常见到的拟合度评估策略除了卡方值与卡方显著性、卡方自由度比两种传统方式之外，还有 CFI 与 RMSEA 指数。有关各项指数运用的时机与判准，列于表 4.2。

表 4.2 各种拟合指数的比较

指标名称与性质	范围	判断值	适用情形
卡方检验			
χ^2 test 理论模型与观察模型的拟合程度	-	$p>.05$	说明模型解释力
χ^2/df (Wheaton et al.) 考虑模型复杂度后的卡方值	-	<2	不受模型复杂度影响
适合度指数			
GFI (Bentler, 1983) 假设模型可以解释观察数据的比例	0~1	$>.90$	说明模型解释力
AGFI (Bentler, 1983) 考虑模型复杂度后的 GFI	0~1*	$>.90$	不受模型复杂度影响
PGFI (Mulaik, 1989) 考虑模型的简效性	0~1	$>.50$	说明模型的简单程度
NFI (Bentler & Bonett, 1980) 比较假设模型与独立模型的卡方差异	0~1	$>.90$	说明模型较虚无模型的改善程度
NNFI (Bentler & Bonett, 1980) 考虑模型复杂度后的 NFI	0~1*	$>.90$	不受模型复杂度影响
替代性指数			
NCP (Bentler, 1988) 假设模型的卡方值距离中央卡方分布的离散程度	-	越接近 0 越好	说明假设模型距离中央卡方的程度
CFI (Bentler, 1988) 假设模型与独立模型的非中央性差异	0~1	$>.95$	说明模型较虚无模型的改善程度特别 适合小样本
RMSEA (Browne & Cudeck, 1993) 比较理论模型与饱和模型的差距	0~1	$<.05$	不受样本数与模型 复杂度影响
AIC (Akaike, 1987) 经过简效调整的模型拟合度的波动性	-	越小越 好	适用于效度复核 非嵌套模型比较
CAIC (Akaike, 1987) 经过简效调整的模型拟合度的波动性	-	越小越 好	适用于效度复核 非嵌套模型比较
CN (Hoelter, 1983) 产生不显著卡方值的样本规模	-	>200	反映样本规模的适 切性
残差分析			
RMR 未标准化假设模型整体残差	-	越小越 好	了解残差特性
SRMR 标准化假设模型整体残差	0~1	$<.08$	了解残差特性

* 指数数值可能会超过范围之外。

（二）模型拟合指数的报告

Hu 与 Bentler (1999) 主张 CFI 与 RMSEA 两个指数都需报告在论文中。尤其是 RMSEA 指数, 当研究者想去估计统计检验力时特别适合。另外, 当研究者想要比较不同的模型, 但是没有嵌套关系时, 则可使用 ECVI、AIC 或 CAIC 指数。

在报告模型拟合度的评鉴数据时, McDonald 与 Ho (2002) 建议必须将测量模型与结构模型的模型拟合度分开列举说明。Anderson 与 Gerbing (1988) 曾经提出一个二阶段估计程序 (two-stage procedure), 他们将结构模型视为测量模型的嵌套模型, 将一个完整的 SEM 模型的测量与结构部分的模型拟合度评鉴区分为两个独立的步骤进行, 如此即符合 McDonald 与 Ho (2002) 的建议。

如果不是使用两阶段估计程序, 在 ML 与 GLS 两种估计程序中, 一个兼含测量模型与结构模型的 SEM 模型整个模型的卡方值可以直接切割成两个独立的卡方值, 一个反映测量模型, 另一个反映结构模型, 然后分别利用相对应的卡方分布进行显著性检验 (Steiger, Shapiro, & Browne, 1985)。

第三节 结 语

本章介绍了多种用来评估 SEM 模型优劣的方法与技术, 并说明如何进行模型修饰的原理与技术。基本上, SEM 模型评鉴的基本概念, 是以参数估计所得到的估计数与实际观察所得到的数据的差异程度作为基础, 透过拟合函数的计算, 进而转换成卡方统计量, 得以进行显著性检验。然而, 受限于卡方检验的诸多问题, 越来越多的研究者改采模型拟合指数评估 SEM 模型的优劣。这些模型拟合指数的选择甚多, 但 McDonald 与 Ho (2002) 指出这些指数的应用也具有有一些限制, 读者应加以留心。

首先, 这些指数尽管都有很明确的意义, 但是从数学或实证的角度来看, 并没有一个强而有力的理论基础支持数字的意义与使用原则, 背后仍存有未知的隐忧。

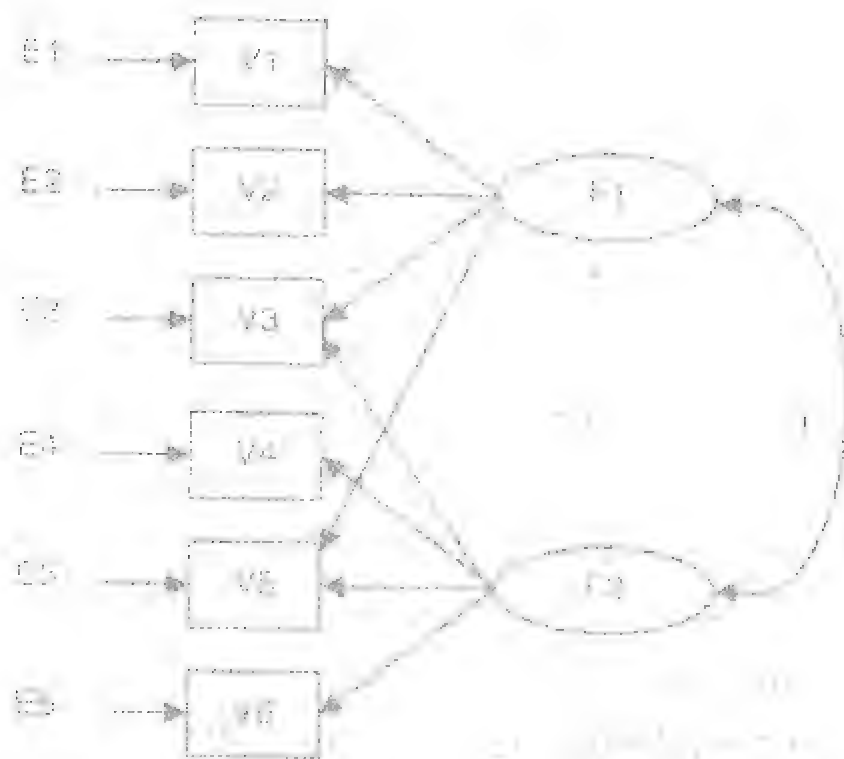
其次，不同的指数的优劣比较仍具有相当争议，尤其是某些指数以独立模型（观察变量间不具有共变关系假设）作为比较基点的做法，其合理性仍有待进一步证明。

第三，SEM 模型的检验应以理论为依归，进行统计决策时，应该具有一定的理论合理性。然而，这些模型拟合指数往往只是反映一种分析技术上的程度，而非理论上的证据。当研究者可以提出无数种可能模型，最佳的指数可能只是反映了某一个模型在技术上的优化，而不是理论上的优化。

最后，不良的模型拟合度往往是因为错误的模式界定所造成，当模型中存在着某些不理想的参数设定时，可能会产生很可观的不拟合度。由于模型拟合指数是一种概括性指数，这些不当的参数设定无法被模型拟合指数来反映。尤其是一些极端明显的不适当参数，如果可以被侦测出来，将可以大幅度影响模型拟合与否的判断，因此 McDonald 与 Ho (2002) 建议 SEM 分析的研究应报告各变量的相关与拟合残差量 (discrepancies)，使得人们可以明确了解各变量的估计状况。

第五章

验证性因素分析与模型修饰



因素分析 (factor analysis) 是一套用来简化变量、分析变量间的群组关系或寻找变量背后共同的潜在构念的统计技术。为了要证实研究者所设计的测验工具的确能够用以测量某一潜在特质的强度与内在结构, 将一组具有共同特性或有特殊结构关系的测量指标, 抽离出背后潜在构念并进行因素关系的探究的统计分析技术, 便是因素分析。

第一节 验证性因素分析原理

一、探索性与验证性因素分析

传统上, 研究者在进行因素分析之前, 并未对于数据的因素结构有任何预期与立场, 而藉由统计量据分析因素的结构, 此种因素分析策略带有浓厚的尝试错误的意味, 因此称为探索性因素分析 (exploratory factor analysis; EFA)。然而, 有时研究者在研究之初既已提出某种特定的结构关系的假设, 例如某一个概念的测量问卷是由数个不同子量表所组成, 此时因素分析可以被用来确认数据的模式是否即为研究者所预期的形式, 此种因素分析称为验证性因素分析 (confirmatory factor analysis; CFA), 具有理论检验与确认的功能。在技术层次来说, CFA 是结构方程模型的一种次模型 (submodel) (Jöreskog & Sörbom, 1993), 除了作为因素结构检证之用, 并可以与其他次模型整合, 成为完整的结构方程模型分析。

探索性因素分析与验证性因素分析最大的不同在于测量的理论架构 (因素结构) 在分析过程中所扮演的角色与检验时机。对 EFA 而言, 测量变量的理论架构是因素分析的产物, 因素结构是从一组独立的测量指标或题目间, 以数学程序与研究者主观判断所决定的一个具有计量合理性与理论適切性的结构, 并以该结构代表所测量的概念内容。换句话说, 理论架构的出现在 EFA 是一个事后 (posterior) 的概念。相对之下, CFA 的进行则必须有特定的理论观点或概念架构作为基础, 然后藉由数学程序确认该理论观点所导出的计量模型是否确实、适当。换句话说, 理论架构对于 CFA 的影响是于分析之前发生的, 计量模型具有理论的先验性, 其作用是一种事前 (priori) 的概念。

在 SEM 架构下, CFA 为一种子模型, CFA 分析的数学原理与统计程序, 都是

SEM 的一种特殊应用。由于 SEM 的模式界定能够处理潜在变量的估计与分析, 具有高度的理论先验性, 因此如果研究者对于潜在变量的内容与性质, 在测量之初即非常明确、详细地加以推演, 或有具体的理论基础, 提出适当的测量变量组成测量模式, 借由 SEM 的分析程序, 可以对于潜在变量的结构或影响关系进行有效的分析。SEM 中对于潜在变量的估计程序, 即是在考验研究者先期提出的因素结构 (测量模型) 的適切性。一旦测量的基础确立了, 潜在变量的因果关联就可以进一步透过多元回归、路径分析的策略 (结构模型) 加以探究。一般而言, CFA 可以说是进行整合性结构方程模型分析的一个前置步骤或基础架构, 亦可以独立进行。

CFA 在结构方程模型整体的发展过程中, 占有相当重要的地位。最初发展的先驱如 Jöreskog 等人, 长期以来积极地改善传统的因素分析的限制, 扩大其应用范围, 最后促成了结构方程模型的出现。在结构方程模型的分析架构中, CFA 所检验的是测量变量与潜在变量的假设关系, 可以说是结构方程模型最基础的测量部分。它不但是结构方程模型中其他后续高等统计检验的基础, 更可以独立应用在信效度的考验与理论有效性的确认 (Bentler, 1989)。

由于 CFA 使用的范围相当广泛, 大大超越了传统 EFA 用来简化数据或抽取因素的单纯目的, CFA 可以用来检验抽象概念或潜在变量的存在与否, 评估测验工具的项目效度与信效度, 并且检验特定理论假设下的因素结构 (参考 Brown, 2006), 因此在 SEM 范式下, CFA 经常被独立使用。然而, 虽然 CFA 可以说是因素分析技术的一大革命, 但是读者仍需注意 EFA 与 CFA 两者的目的不同, 使用的时机也不一样。从研究的立场来看, CFA 并不足以完全取代 EFA, 两者反而具有相辅相成的功效, 因此建议读者两者均要熟悉。

二、潜在变量的因素分析

以 SEM 的术语来看, 潜在变量的因素分析就是所谓的 CFA, 可以说是心理计量学者在处理潜在特质测量问题时, 最具有创意的表现, 使得过去传统因素分析与路径分析限制得以突破并加以融合。潜在变量反映的是研究者关心但无法直接观察与测量的现象或假设性构念。从外表来看, 构念所反映的通常只是一组有相互关联的行为现象, 经由研究者基于理论的推导、文献的支持或个人主观的分析与演绎, 所提出的理论上的存在假设 (Nunnally & Bernstein, 1994)。

为了使构念得以有效的测定,需要经过下列四个程序:第一,构念必须要有明确的操作型定义界定其内容与范畴。第二,用以测量构念的指标能够被明确地指出。第三,测量同一构念的指标必须具有相当的一致性,测量不同构念的指标则有相当的区辨性,多元指标的一致与区辨性应能从观察资料中检验得出。第四,经由统计检证的程序,观察数据可以用以支持或推翻构念是否存在的假设。上述程序称之为构念有效化 (construct validation) (Cronbach & Meehl, 1955)。从 SEM 的术语来说,这个过程就是检验构念的测量模型的过程。当从测量变量所搜集到的共变量矩阵与测量模型所推导、计算出来的共变量矩阵加以比对,两者若相符达到一定的程度,我们即无法推翻测量模型存在的可能性,也就是研究者所提出的潜在构念的概念建构获得支持。

Kline (1998) 指出测量模型的检验具有两个值得注意的逻辑上错误:第一,当测量数据无法指出测量模型的不適切时(即无法推翻测量模型存在的可能性),并不能说明测量模型所代表的抽象建构确实存在。因为同一组测量指标有多种不同的组合方式,也就是说可能有其他替代模型存在的可能。对于特定一个理论模型与观察数据的共变结构的拟合比对,其数据所反映的是一种数学的现象,倘若研究者提出另一套替代的测量模型与相同的观察数据相比对,也可能获得支持的结论。虽然 SEM 提供一套竞争模型比较的策略决定一个最佳的模型,但是基于 SEM 的数学程序,此一逻辑性错误的威胁始终存在。此外,另一个原因是测量模型所代表的构念究竟为何也无法从 SEM 的检证程序得以确认。由研究者根据测量的指标与概念发展的过程以人为方式为该构念命名,背后可能的错误称为命名陷阱 (naming fallacy)。命名陷阱一直威胁着因素分析的应用,研究者不宜单凭测量模型的支持推导某特定构念确实存在的结论。

第二种逻辑上的错误是抽象概念实质化 (reification) (Gould, 1981)。当 SEM 程序确认某一特定的测量模型最能反映观察数据时,仅代表某种抽象概念的测量程序的有效性被支持,但是该抽象建构是否事实存在,则无法确知。也就是说,假设性的构念不一定反映真实世界的现象,如果研究以测量模型的结果去描绘客观的事实,则犯了抽象概念实质化的逻辑错误。

三、验证性因素分析的特性

在 SEM 的术语中, 测量模型的检验程序称之为确认性因素分析, 图 5.1 即是一个典型的 CFA 测量模型。图 5.1 中有两个具有相关的潜在变量 F1 与 F2, F1 由 V1 至 V3 三个指标来测量, F2 由 V4 至 V6 三个指标来测量, E1 至 E6 分别代表六个测量变量的测量误差。从潜在变量指到测量变量的单箭头, 代表研究者所假设的潜在变量对于测量变量的直接因果关系。经由统计过程对于这些因果关系的估计数称为因素载荷, 有标准化与未标准化两种形式, 性质类似于回归系数。

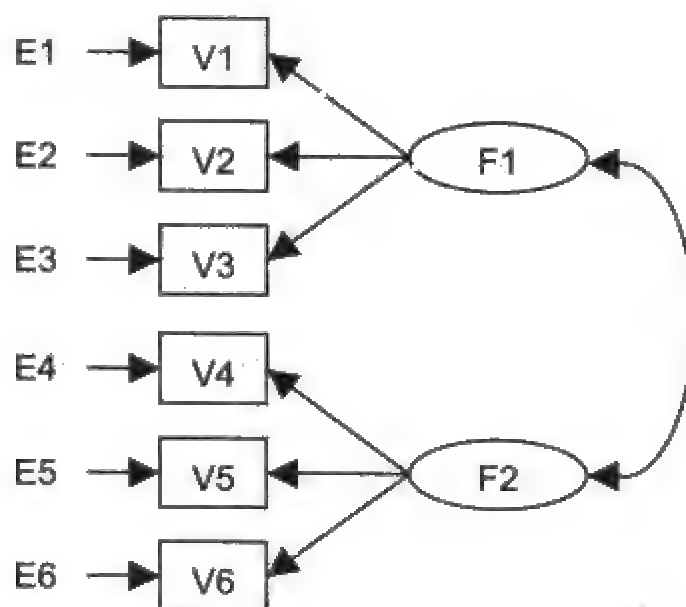


图 5.1 典型的 CFA 测量模型图

在整个模型当中, 研究者所能具体测量的是六个测量变量, 其背后受到某些共同的潜在变量的影响, 因此测量变量可以说是内生变量, 潜在变量与测量误差则为外源变量。从变异量的拆解原理来分析, 每个测量变量的变异量可以被拆解成两部分: 共同变异 (common variance) 与独特 (或误差) 变异 (unique variance)。以 F1 为例, V1、V2、V3 是用来测量潜在特质 F1 的指标或测量题目, 其背后受到同一个潜在因素的影响。从数学关系来说, 即是三个变量共变的部分, 而测量误差的部分就是三个测量变量无法被该潜在变量解释的独特变异量 (uniqueness) 或残差 (disturbance), 彼此相互独立。

四、测量误差与方法效应

CFA 测量模型分析与路径分析最大的不同在于误差的处理。传统路径分析当中的每一个变量被视为可直接测量、可信赖的变量，误差的概念仅反映在外源变量对于内生变量无法预测与解释的部分。虽然研究者不会忽略外源变量与内生变量本身可能存在着测量误差，但是测量误差的估计则是独立于路径模型之外来进行的，即以信度估计的方式表达每一个变量的可信度。路径分析的本身却是假设变量是无误差的测量，这个假设显然是很容易被违反的，尤其当变量本身带有高度的抽象特质或是当变量分数的信度偏低之时。

CFA 测量模型能够有效处理路径分析当中变量具有一定测量误差的问题。CFA 测量模式不但将每一个具有抽象特性的外源或内生变量以潜在变量的形式表示，同时对于潜在变量的每一个测量变量进行误差的估计，即独特变异量的分析。从测量变量所拆解得出的独特变异量可能包含两种类型的测量误差。第一是随机误差 (random error)，也就是传统信度估计所处理的测量误差。造成随机误差的原因很多，例如测量的过程因素、受试者个人因素、工具的因素等。但这些因素对于测量分数并无系统化的影响（如系统性的高估或低估），因此称为随机性的误差来源。第二则是系统误差 (systematic error)，对于测量分数会有系统化的影响，使测量分数以特定的模式偏离实际的真分数。理论上，系统化误差可以从测量变量中的独特变异量中抽离出来，因为系统性误差的存在，是在潜在变量之外仍有其他影响分数的变异来源导致的。最明显的一个例子是方法效应 (method effect)，也就是研究过程中对于变量的测量基于某一种特殊的方法，例如纸笔测验，造成测量分数的系统性变化，无法反映真分数的一种现象 (Marsh, 1988; Kenny & Kashy, 1992)。

在测量上，方法效应对于测量分数的正确运用有非常大的影响，因为方法效应会使得测量分数当中除了由于测量的特质引起的变异 (trait variance)，增加了另一种因为工具或方法所产生的方法变异 (method variance) (Bagozzi, 1993)，而混淆了研究数据的意义。在相关系数的计算上，使用同一种工具所测得的分数的相关系数，很可能因为使用了共同的方法而造成扩大与膨胀 (Mitchell, 1985)。针对这些基于工具引发的系统误差，CFA 测量模型提供了技术上的解决策略，藉由共变关系的分析，配合多维测量的假设（见下一节），可以有效地估计存在于独特变异当中的系统性方法变异 (Marsh, 1989; Marsh & Bailey, 1991; Kenny & Kashy, 1992)。此外，研究者发展出 CFA 多重特质多重方法矩阵法 (multitrait-multimethod matrix)

程序来分析多重方法所造成的多重方法效应的问题 (Chiou & Hocevar, 1995, Chiou, 1995; Lowe & Ryan-Wenger, 1992; Marsh & Hocevar, 1988)。在文献上, CFA 测量模式的分析对于系统误差的处理能力受到学者相当的重视, 对于解决计量研究者长期以来的困扰有其一定的贡献 (见 Fisicaro, 1988, Ilgen, Barnes-Farrell, & McKellin, 1993; Murphy, Jako, & Anhalt, 1993; Podsakoff & Organ, 1986)。

五、单维测量与多维测量

CFA 测量模式与传统因素分析一个相当大的不同, 在于测量变量与潜在因素之间的组合形态, 不但可以由研究者依其假设来指定, 更可以突破单一一个变量只能用以反映单一一个潜在变量, 即单维测量 (unidimensional measurement) 的限制。

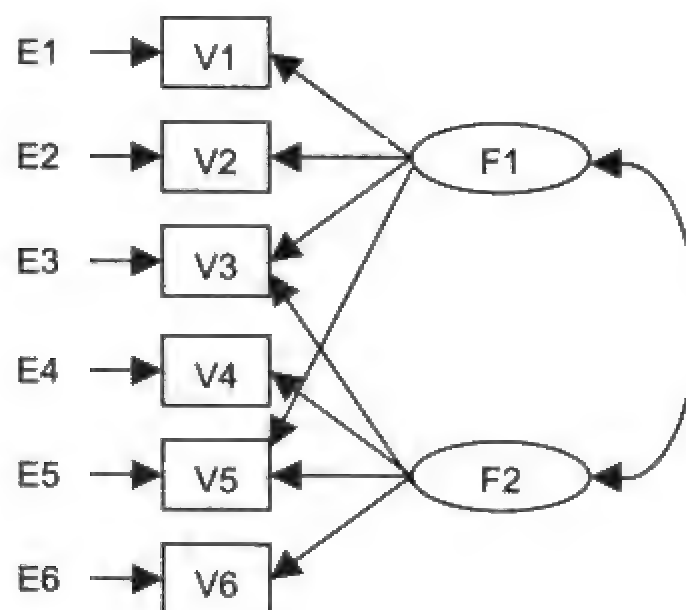


图 5.2 多维测量的 CFA 测量模型图

以图 5.2 为例, 每一个测量变量皆与一个特定的潜在变量相联结, 但是 V3 与 V5 则另外尚有额外的联结 (V3←F2、V5←F1), 也就是说 V3 与 V5 同时与两个潜在变量有关, 反映出 V3 与 V5 两个变量的测量有其他变异来源的特殊测量模型, 称为多维测量 (multidimensional measurement)。

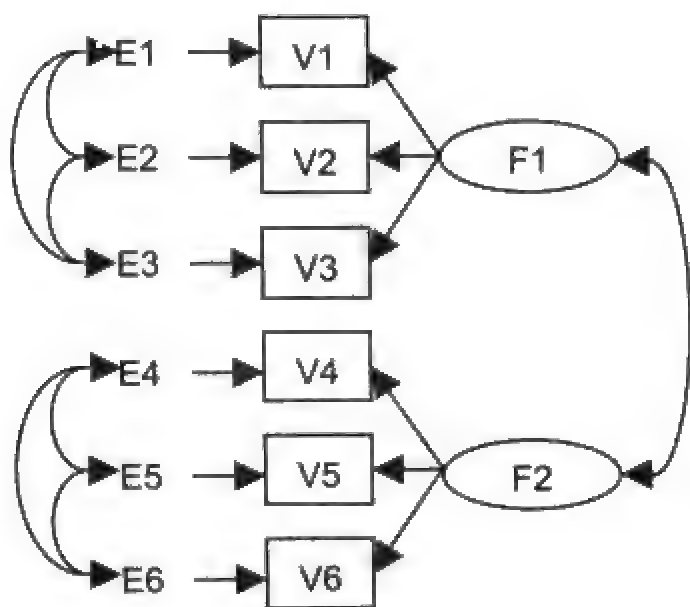
CFA 测量模式除了允许测量变量与潜在变量可以具有多维关系之外, 测量变量

的误差项（独特变异）也可以与其他变量存在假设性的共变，也就是说，CFA 测量模型在技术上允许测量变量的误差项为多维的测量。最普遍的一种现象是使用相关误差（correlated measurement error）的测量模型，当误差项存在有意义的相关，代表测量变量除了受到特定潜在特质的影响之外，尚有其他未知的影响来源，而须由误差项的共变分析来估计之。图 5.3 描绘出两种不同的相关误差测量模型。

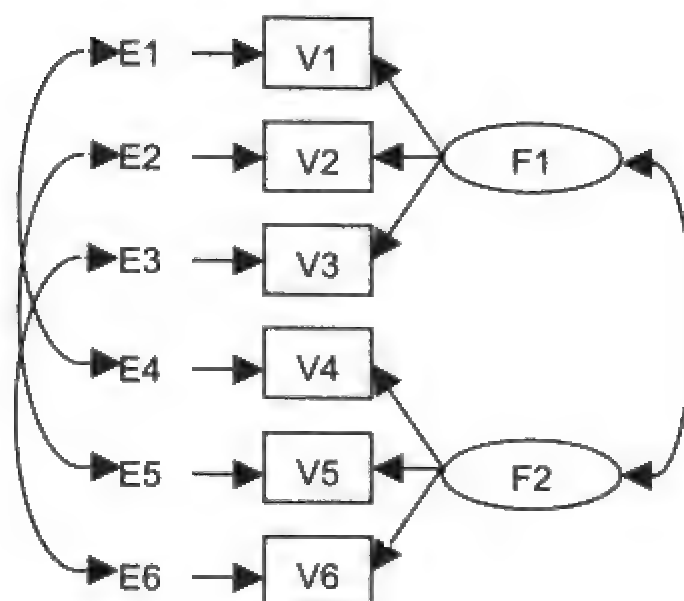
图 5.3 (a) 与图 5.3 (b) 中，误差项之间具有假设存在的共变，图 5.3 (a) 的关联误差发生在同一个潜在因素内，称为因素内关联误差模型，而图 5.3 (b) 的关联误差则是跨越了不同的因素，称为因素间关联误差模型。

因素内关联误差模型最典型的例子即是前一节所提及的方法效应，也就是说测量变量都是由同一种测量工具所测量，例如 F1 是以三个 Likert 式自陈量表题目 V1、V2、V3 来测定，F2 由另一种方法来测定。对于同一个因素（如 F1），三个测量变量除了受到该潜在因素的影响外，还受到方法效应的影响，反映在误差项的相关上（ $E1 \leftrightarrow E2$ 、 $E1 \leftrightarrow E3$ 、 $E2 \leftrightarrow E3$ ）。

因素间关联误差模型最典型的例子是再测信度的测量。假设 F1 因素与 F2 因素分别代表在两个时间点下同一个潜在特质的两次测量，V1、V2、V3 与 V4、V5、V6 是同一组题目，即 $V1=V4$ 、 $V2=V5$ 、 $V3=V6$ 。时间变动将造成测量分数的波动，三个题目在两次测量中的误差项的两两相关（ $E1 \leftrightarrow E4$ 、 $E2 \leftrightarrow E5$ 、 $E3 \leftrightarrow E6$ ）反映了特定题目因为时间变动的波动情形。



(a) 因素内关联误差 CFA 模型



(b) 因素间关联误差 CFA 模型

图 5.3 两种关联误差的 CFA 测量模型图

第二节 测量模型的内部拟合检验

前面已经讨论了测量模型的基本原理与模型的可能变形。然而一个测量模型可否被接受，以及参数估计的优劣好坏，除了从模型的整体拟合来看之外，还必须从模型的内在质量来衡量每一个潜在变量的適切性，又称为内部拟合。这就好比过去我们以方差分析 (Analysis of Variance, ANOVA) 来进行多个平均数的差异比较时，首先要以 F 检验来考验平均数之间是否存在一个整体性、有意义的差异，称为整体检验 (overall test)，然后再去检验差异究竟发生在哪里，称为事后检验 (Post-hoc test)。多元回归分析也遵循相同的法则，亦即先对整体回归模型的解释力 (R^2) 以 F 检验进行考验。一旦整体模型的回归解释力具有统计显著性之后，才进行个别解释变量的参数检验。

结构方程模型虽然没有一个单一的解释力 (R^2) 量数得以进行模型的整体检验，但是可以就假设模型与观察模型之间的拟合程度进行整体检验。当模型拟合性被接

受之后,我们得以针对个别的因素质量进行检测。在本质上,模型拟合性检验就是一种整体检验,个别因素的检验就是一种事后的评估程序。Hair等人(2006)认为,在验证性因素分析中,除了报告模型拟合指数之外,还必须进一步了解测量模型当中的个别参数是否理想(项目信效度),各潜在变量的组合情形是否稳定可靠(构念的信效度)。如果某些参数不甚理想,可以藉由模型修饰程序去除不良题项或增加参数提高模型的内在拟合。在具体做法上,比较为多数人所采用的策略包括四项检验:项目质量、组合信度(ρ_c)、平均变异萃取量(ρ_v)、因素区辨力,以下将分别介绍之。

一、项目质量检验

Bollen(1989)指出,构成一个有意义的潜在变量的最重要前提,是一组能够反映潜在构念意义的观察指标。换言之,构成潜在变量的题目必须具有相当的信效度,否则无法支撑一个潜在变量模式。

在古典测验理论中,信度(reliability)代表测量的可靠程度(trustworthiness),或不受测量误差影响真分数测量的程度。效度(validity)则反映了测量工具能够正确无误地测出潜在特质的程度,也就是研究者可以掌握到抽象意义的程度。两者必须同时兼具,才能确保测量的质量。

就组成一个因素的个别题目来说,当题项测量误差越小,表示测量题目受到误差的影响越小,能够测到真分数的程度越高。SEM的测量模式中,测量残差由 Θ_δ 与 Θ_ε 两个矩阵的对角线的参数(δ 与 ε 系数)反映,而这两组参数与因素载荷具有函数关系,当因素载荷越高,测量残差越小。Bagozzi与Yi(1988)认为测量模型当中的测量残差必须具有统计显著性,才能确立一个潜在变量是由一组带有测量误差的观察变量所形成的这个前提。相反地,如果测量误差太微弱而未达统计显著性(或因素载荷太高,超过.95时),则意味该题足以完全反映该潜在构念的内容,测量模型的合理性即不复存在(Anderson & Gerbing, 1988)。除此之外,因素载荷的系数正负号也应符合理论预期,更不应出现超过 ± 1 的数值。这些条件的维系,是测量模型的基本拟合指数(preliminary fit criteria)(Bagozzi & Yi, 1988)。

延续前述的讨论,因素载荷一方面除了反映测量误差的影响之外,也同时反映了个别题目能够用来反映潜在变量的程度。Hair等人(2006)认为一个足够大的因

素载荷，代表题项具有良好的聚敛效度 (construct validity)。一般而言，当因素载荷大于.71 时，即可以宣称项目具有理想质量，因为此时潜在变量能够解释观察变量将近 50% 的变异，这个 $\lambda \geq .71$ 指标可以说是基本拟合指数当中最明确的一个判准 (Bagozzi & Yi, 1988; Hair et al, 2006)。

事实上， $\lambda \geq .71$ 原则其实来自于传统因素分析当中共同性的估计 (communality)，亦即个别题目能够反映潜在变量的能力指标。Tabachnicka 与 Fidell (2007) 具体提出了下列标准 (见表 5.1)：当因素载荷大于.71，也就是该因素可以解释观察变量 50% 的变异量之时，是非常理想的状况；当载荷大于.63，也就是该因素可以解释观察变量 40% 的变异量之时，是非常好的状况；但若载荷小于.32，也就是该因素解释不到 10% 的观察变量变异量，是非常不理想的状况。通常这类题目虽然是形成某个因素的题项，但是贡献非常小，可以考虑删除该题，以提高整个因素的一致性。

表 5.1 因素载荷的判断标准

λ	λ^2	状况
.71	50%	优秀
.63	40%	非常好
.55	30%	好
.45	20%	普通
.32	10%	不好
.32 以下		不及格

一般来说，社会科学研究者所编制的量表的因素载荷都不会太高，这可能是受限于测量本质的特性 (例如态度测量的范围太广不易聚焦、构念过于模糊不易界定)、外在干扰与测量误差的影响、甚至于构念本质是形成性或反映性等争议。此时，建议采用 Tabachnicka 与 Fidell (2007) 所建议的标准 (例如 $\lambda \geq .55$ 即可宣称良好)，而不必坚守 $\lambda \geq .71$ 原则。

SEM 分析软件中的多元相关平方 (squared multiple correlation, SMC) 数据，反映了个别测量变量受到潜在变量影响的程度。SMC 越高，表示真分数所占的比重越高；相对地，SMC 越低，表示真分数所占的比重越低，信度越低。SMC 的计算原理如下：

$$SMC_{\text{vari}} = \frac{\lambda_i^2}{\lambda_i^2 + \Theta_{ii}} \quad (5.1)$$

其中 λ_i 为个别测量变量的因素载荷,取平方后除以总变异量(解释变异量加误差变异量),即为个别题目的信度估计数。值得注意的是, SMC 的计算是以单维假设为基础的信度估计数,也就是一个测量变量仅受到单一一个潜在变量的影响(单一真分数变异来源)。如果一个测量变量受到两个或以上的潜在变量的影响,该式即不适用。

二、组合信度 (ρ_c)

(一) 测量信度的概念

传统的测验观点认为,对于被测量的心理特质,在一定的时间范围内,每个人都具有一个特定的水平或强度。测验的主要目的,就是利用一组题目去评估个体在这个特质上的水平或强度。如果一组题目所构成的测验真的可以测到这个人“真实”的特质强度,反映在测验得分上,称为真实分数(true scores)。这个真实分数是一个理论上存在的分数,代表受测者的实际心理特质内涵与真实的心理运作历程。测量分数当中真实分数所占的比例越大,表示测量信度越高。在测验实务上,准确测得人类心理特质的真实分数是所有测验的终极目标。从这个目标所延伸出来的测验理论,一般即称为古典测验理论(classical test theory)。

古典测验理论最关心的是真分数如何被测量到,但由于真分数无法得知,因此仅能从另一个方向去思考,亦即影响或干扰真分数测量的因素越少,真分数被测得的可能性越高。因此得到一个基本假设为:观察分数=真实分数+误差分数。当误差为0,观察分数可以完全反映真实分数。若以方差的概念来表示,观察分数的变异可以 σ_{total}^2 表示;而真实分数的方差可以 σ_{true}^2 表示;误差分数的方差则反映了测量误差波动情形,以 σ_{error}^2 表示。三者关系如下:

$$\sigma_{\text{total}}^2 = \sigma_{\text{true}}^2 + \sigma_{\text{error}}^2 \quad (5.2)$$

或

$$1 = \frac{\sigma_{\text{true}}^2}{\sigma_{\text{total}}^2} + \frac{\sigma_{\text{error}}^2}{\sigma_{\text{total}}^2} \quad (5.3)$$

公式(5.3)中, 真实分数的变异除以变量的总变异量, 代表一个测量分数是否具有信度的指标, 即为信度系数(coefficient of reliability), 以 ρ 表示(Lord & Novick, 1968)。如公式(5.4)所示:

$$\rho = \frac{\sigma_{\text{true}}^2}{\sigma_{\text{total}}^2} = 1 - \frac{\sigma_{\text{error}}^2}{\sigma_{\text{total}}^2} \quad (5.4)$$

(二) SEM 的组合信度估计

SEM测量模式的信度估计基本上延续了古典测量理论的观点。将信度视为真实分数所占的比例而测量误差的变异即为观察分数当中无法反映真实分数的残差变异量。对于个别测量题目来说, 由于测量变量分数的变动会受到潜在因素与测量误差的影响, 而潜在因素所影响产生的变异即代表真实分数的存在, 因此, 信度可以用测量变量的变异量能够被潜在变量解释的百分比(proportion of variance of a measured variable)来表示。Fornell与Larker(1981)基于前述SMC的概念, 提出了一个非常类似于内部一致性信度系数(Cronbach's α)的潜在变量的组合信度(composite reliability; CR或 ρ_c):

$$CR = \rho_c = \frac{(\sum \lambda_i)^2}{[(\sum \lambda_i)^2 + \sum \Theta_{ii}]} \quad (5.5)$$

上式中, $(\sum \lambda_i)^2$ 为因素载荷加总后取平方之数值, $\sum \Theta_{ii}$ 为各观察变量残差方差的总和。当测量模式中带有残差相关时, 残差变异量估计数会因为残差间的相关而降低, 因此 ρ_c 的估计必须将残差相关纳入计算(Raykov, 2004; Brown, 2006), 公式如下:

$$\rho_c = \frac{(\sum \lambda_i)^2}{[(\sum \lambda_i)^2 + \sum \Theta_{ii} + 2\sum \Theta_{ij}]} \quad (5.6)$$

其中 $\sum \Theta_{ij}$ 为第*i*与第*j*题残差共变的总和。换言之，利用SEM来估计模型之余，尚可以进行测量工具的信度的估计。而且，除了整体量表的信度得以估计之外，也可以计算个别测量题目的信度。唯一的缺点是在计算整个因素（量表）的信度之时，必须以人为的方式来计算，LISREL尚无模块得以自动产生。

若要利用 LISREL 来进行计算，Brown (2006) 提供了一组语法可用来产生组合信度值，其中利用非线性限定参数可以产生一组虚拟潜在变量来估计三个数据：真分数变异、总变异以及两者的比值。有兴趣的读者可以参考书中的范例。

依据古典测验理论的观点，量表信度需达.70 才属比较稳定的测量，在SEM的测量模式也多沿用此一标准。但此标准的达成必须使各题的因素载荷平均达.70 以上，社会科学领域的量表不易达到此水平，因此Bagozzi与Yi (1988) 建议 ρ_c 达.60 即可。Raine-Eudy (2000) 的研究指出，组合信度达.50 时，测量工具在反映真分数时即可获得基本的稳定性。

（三）组合信度的区间估计

前述的 ρ_c 估计是一种点估计，如果考虑模型估计的随机性可能造成的误差影响，信度估计也应把样本波动因素纳入考虑，采取区间估计的做法报告因素的信度，说明我们的测量模式估计得到的 ρ_c 可以反映真正的总体信度的95%或99%的置信区间为何。

进行区间估计最重要的步骤是计算估计数的抽样标准误（standard error of sampling），Raykov (2002) 提出验证性因素分析的因素信度的标准误计算法则，

$$SE(\rho_c) = \sqrt{D_1^2 Var(u) + D_2^2 Var(v) + 2D_1 D_2 Cov(u, v)} \quad (5.7)$$

其中u是标准化因素载荷估计数的总和，v是误差方差估计数的总和， $Cov(u, v)$ 则是两者的协方差， D_1 与 D_2 是量表信度系数分别对u与v作偏微分，得到估计数：

$$D_1 = \frac{2uv}{(u^2 + v)^2} \quad (5.8)$$

$$D_2 = \frac{u^2}{(u^2 + v)^2} \quad (5.9)$$

一旦获得标准误的估计数后, 即可获得信度的 95% 或 99% 的置信区间:

$$95\%CI(\rho_c) = \rho \pm 1.96SE(\rho_c) \quad (5.10)$$

$$99\%CI(\rho_c) = \rho \pm 2.59SE(\rho_c) \quad (5.11)$$

三、平均变异萃取量 (ρ_v)

先前已经提及, 测量题目的因素载荷越高, 表示题目能够反映潜在变量的能力越高, 因素能够解释各观察变量的变异的程度越大。因而可以计算出一个平均变异萃取量 (Average Variance Extracted, AVE 或 ρ_v), 来反映一个潜在变量能被一组观察变量有效估计的聚敛程度指标 (Fornell & Larker, 1981)。公式如下:

$$AVE = \rho_v = \frac{\sum \lambda_i^2}{(\sum \lambda_i^2 + \sum \Theta_{ii})} \quad (5.12)$$

上式中, 分母为各题的因素载荷平方加上误差变异, 相加为 1。因此分母即为题数 n :

$$\rho_v = \frac{\sum \lambda_i^2}{n} \quad (5.13)$$

换言之， ρ_v 指标就是各因素的各题因素载荷平方的平均值。如果配合前述的 $\lambda \geq 71$ 原则，那么 ρ_v 的判准也即是.50 (Anderson & Gerbing, 1988, Hair et al., 2006)。当 ρ_v 大于.50，表示潜在变量的聚敛能力十分理想，具有良好的操作型定义化 (operationalization)。

从数学过程来看， ρ_v 的概念其实就是传统探索性因素分析当中的特征值 (eigenvalue)。亦即当各观察变量提供一个单位变异量时，各因素的解释变异量，也就是潜在变量变异量占总变异的百分比。换言之，验证性因素分析当中的每一个因素，就是执行一次单因素的探索性因素分析的结果， ρ_v 即为该单一因素的特征值。因此， ρ_v 的解释宜以概念的本身来解释，而不宜解释成聚敛效度 (Hair et al., 2006)。

四、因素区辨力

Hair等人 (2006) 除了引用 ρ_v 作为聚敛能力的指标，也指出了验证性因素分析估计结果所得到的潜在变量必须具有区分效度 (discriminant validity)，亦即不同的构念之间必须能够有效分离。

在具体的 CFA 操作技术上，有三种方式可以用来检验潜在变量的区辨力。第一种是相关系数的区间估计法，如果两个潜在变量的相关系数的 95%置信区间涵盖了 1.00，表示构念缺乏区辨力。

第二种方法是竞争模式比较法，利用两个 CFA 模型进行竞争比较。一个 CFA 模型是令两个构念之间相关自由估计 (效度模型)，另一个 CFA 模型则是将相关设为 1.00 (完全相关模型，此模型也即等同于单一因素模型)，完全相关模型由于少一个有待估计的参数，自由度多 1，模式的拟合度也会较低。如果效度模型没有显著地优于完全相关模型，即代表两个构念间缺乏区辨力 (Anderson & Gerbing, 1988, Bagozzi & Phillip, 1982)。

第三种方法是平均变异萃取量比较法，比较两个潜在变量的 ρ_v 平均值是否大于两个潜在变量的相关系数平方 (Fornell & Larcker, 1981)。

Hair 等人 (2006) 将这些测量模型的内在质量的各种要求加以整理后，认为这些检测都要符合的情况下，测量的构念效度即可获得确保。但是此说法应审慎为之，因为测量模型的内在拟合理想或许可以提供聚敛与区分效度的证据 (Anderson &

Gerbing, 1988), 但非构念效度的充分条件。

在心理计量领域, 效度是一个建构的过程, 而非可以从单一统计量数来获取充分支持 (Messick, 1989)。以上述这些程序检验所获得的证据或许可以作为测量模式的质量评估部分证据与参考, 但是要作为测量工具能反映构念效度的充分证据还有一段距离。

更进一步地, 测量工具的构念效度无法藉由待测量表的本身来自我证实。换言之, 抽象特质是否能被正确地测量得到的计量证据, 必须超越研究者所关心的测量工具本身, 以其他方式的测量工具求得合理的关系, 或以实验手段证实测量分数的有效性 (Meehl, 1959; Anastasi & Urbina, 1997)。

另一个值得注意的是, Hair 等人所谓的聚敛效度与传统的定义有所不同。Campbell 与 Fiske (1959) 认为聚敛效度是指以“不同方法”测量相同特质的相关要高于所有的相关。此时, 不同方法是指不同的量表或不同的测量方式 (例如自评与他评), 而非一个分量表当中的不同题目 (不同题目测量相同特质的相关或因素载荷反映的是测量信度)。进一步地, 区分效度的达成是在当不同方法或相同方法测量不同特质的相关要高于不同方法测量相同特质的相关, 此观点也无法单纯从单一一个量表的检验中获得。

因此, 本节虽然引述了 Hair 等人 (2006) 诸多的策略与观点, 说明了个别题目质量、 ρ_c 、 ρ_v 、因素区辨力的做法, 但是在此必须提醒读者在论及测量的“构念效度”时, 应采取更审慎保守的态度。尤其应避免将 ρ_v 视为聚敛效度的唯一证据。

一般现行的验证性因素分析或结构方程模型分析研究, 均把因素载荷与测量误差当作测量稳定与一致性的指标, 亦即信度的概念, 而 ρ_c 与 ρ_v 也反映了潜在变量的整体稳定可靠性, 也是信度的概念。至于透过验证性因素分析方法所获得的理想模型, 多以测量工具具有因素效度 (factorial validity) 来相称 (Byrne, 1994; Bentler, 1995), 而因素效度仅是构念效度的一部分 (Anastasi & Urbina, 1997)。这就好像传统进行探索性因素分析时, 我们所得到的因素结构只能作为量表因素效度的证据。至于构念效度, 还需要以其他方式从待检工具以外的途径来举证 (例如平行测量的相关)。

第三节 验证性因素分析的实例

基本上, CFA 可以视为 SEM 当中的测量模型检验。当一个 SEM 模型中, 如果仅涉及测量模型的检验, 而没有结构模型的概念, 即是验证性因素分析。在此种模型当中, SEM 所处理的仅是测量变量与潜在变量的关系。

CFA 虽然仅处理测量变量与潜在变量的关系的检验, 但是在实际研究上, 即可以应用在两种情况下。第一是应用在测量工具发展时, 用以评估测量工具的因素结构是否恰当。第二是探讨潜在变量之间的关系, 是否与特定的理论观点相符, 称为理论概念的检验。以下, 我们便以一个实际的范例说明 CFA 应用于测量工具检验的分析技术。

一、验证性因素分析的操作步骤

验证性因素分析的执行可以分成几个步骤。第一个步骤是发展假设模型(hypothetic model)。也就是针对测量的题目的潜在结构关系, 基于特定的理论基础或是先期的假设, 提出一个有待检验的因素结构模型。从结构方程模型的术语来说, 就是要建立一套假设的测量模型。

第二个步骤是进行模型的识别, 也就是将研究者所欲检验的测量模型, 转换成符合 SEM 分析的模型, 以便利用统计软件进行分析。此一步骤的完成必须非常谨慎地计算模型的识别性, 以避免 SEM 执行失败。除此之外, 研究者必须熟习 SEM 分析软件的程序语言或操作方式, 将所欲检验的模型写入 SEM 分析软件的指令系统中。

步骤三是执行 SEM 分析, 进行参数估计与模型检验。其中最重要的一个工作是将研究数据整理出适合 SEM 分析的数据库类型。一般而言, 使用原始数据虽然较为方便, 但是由于原始数据库当中, 往往包含有许多与因素分析无关的变量, 以及许多遗漏值存在。因此, 如果要使用原始数据库, 应先清理数据库。比较简单的方法则是直接取用测量变量的相关或共变量矩阵。

第四个步骤是进行结果分析, 也就是分析 SEM 分析的报表结果, 检验各项数据的正确性。第五个步骤是模型的修正, 以获得较佳的数据与结果。

在验证性因素分析当中，模型修饰多从三个方向进行。第一是各题目与潜在变量（因素）之间的关系的确认。检测个别的题目是否如同假设般受到特定因素的影响，或者是受到其他因素的影响，甚至于是否受到多个因素共同的影响。第二是从测量残差的修正着手进行模型修饰，也就是利用测量误差间相关性检视测量模型的拟合度。第三是从因素间的相关情形检验整理测量模型的拟合度。

最后一个步骤是完成 SEM 分析，并做出报告。此时，SEM 模型的分析已经完成，研究者必须从繁杂的统计报表中找出重要的参数数据，加以整理并得到最后的分析结论。

二、验证性因素分析的操作

（一）模型界定

本范例是以组织创新气氛的测量为例，进行测量工具发展的验证性因素分析的操作示范。为了便于说明，本范例仅取用《组织创新气氛量表》（邱皓政，1999）的 18 题短题本进行操作示范。样本是 384 位来自于台湾某家企业的员工。量表题目与描述统计量如表 5.2。

组织创新气氛量表为 Likert 式六点量尺自陈量表，用以测量组织成员对于组织创新气氛的知觉。该量表题目的编写，系由研究者所执行的前导研究，发现影响组织创新气氛知觉的因素，包括“组织价值”、“工作方式”、“团队合作”、“领导风格”、“学习成长”、“环境气氛”等六个因素，因此针对这些因素编写题目，发展出评定量表。

在本范例中，每一个因素仅取出三个题目作为代表，因此共有 18 个题目（代号 A1 至 F3）。受测者在这些题目的得分越高，代表所知觉到的组织气氛越有利于组织成员进行创新的表现。由于部分成员在部分题目上表示无法填答的情形，因此，实际应用于 SEM 分析的样本为完全作答的 350 位。这个 18 题的评定量表，基于研究者所提出的先期结构（六因素测量模型），六个因素与 18 个测量变量的关系可以利用图 5.4 的假设模型来表示。

表 5.2 组织创新气氛量表 18 题短题本描述统计量

题 目	Mean	SD
1. 我们公司重视人力资产、鼓励创新思考。	4.42	0.98
2. 我们公司下情上达、意见交流沟通顺畅。	4.31	1.02
3. 我们公司能够提供诱因鼓励创新的构想。	4.07	0.97
4. 当我有需要,我可以不受干扰地独立工作。	4.02	1.16
5. 我的工作内容有我可以自由发挥与挥洒的空间。	4.25	1.16
6. 我可以自由地设定我的工作目标与进度。	4.24	1.09
7. 我的工作伙伴与团队成员具有良好的共识。	4.37	0.98
8. 我的工作伙伴与团队成员能够相互支持与协助。	4.34	1.03
9. 我的工作伙伴与团队成员能以沟通协调来化解问题与冲突。	4.31	1.05
10.我的主管能够尊重与支持我在工作上的创意。	4.83	0.94
11.我的主管拥有良好的沟通协调能力。	4.95	0.84
12.我的主管能够信任部属、适当地授权。	4.83	0.91
13.我的公司提供充分的进修机会、鼓励参与学习活动。	4.63	0.97
14.人员的教育训练是我们公司的重要工作。	4.73	1.01
15.我的公司重视信息收集与新知的获得与交流。	4.70	0.98
16.我的工作空间气氛和谐良好、令人心情愉快。	4.23	1.17
17.我有一个舒适自由、令我感到满意的工作空间。	4.63	1.09
18.我的工作环境可以使我更有创意的灵感与启发。	4.49	0.94

由于组织气氛的知觉是一个多向度的个人知觉现象,以 18 个测量变量测量受测者的知觉强度,同时不同的知觉向度(六个因素)之间可能具有相关。根据这些条件,模型界定情形说明如下,并以路径图标示各参数(图 5.5)。

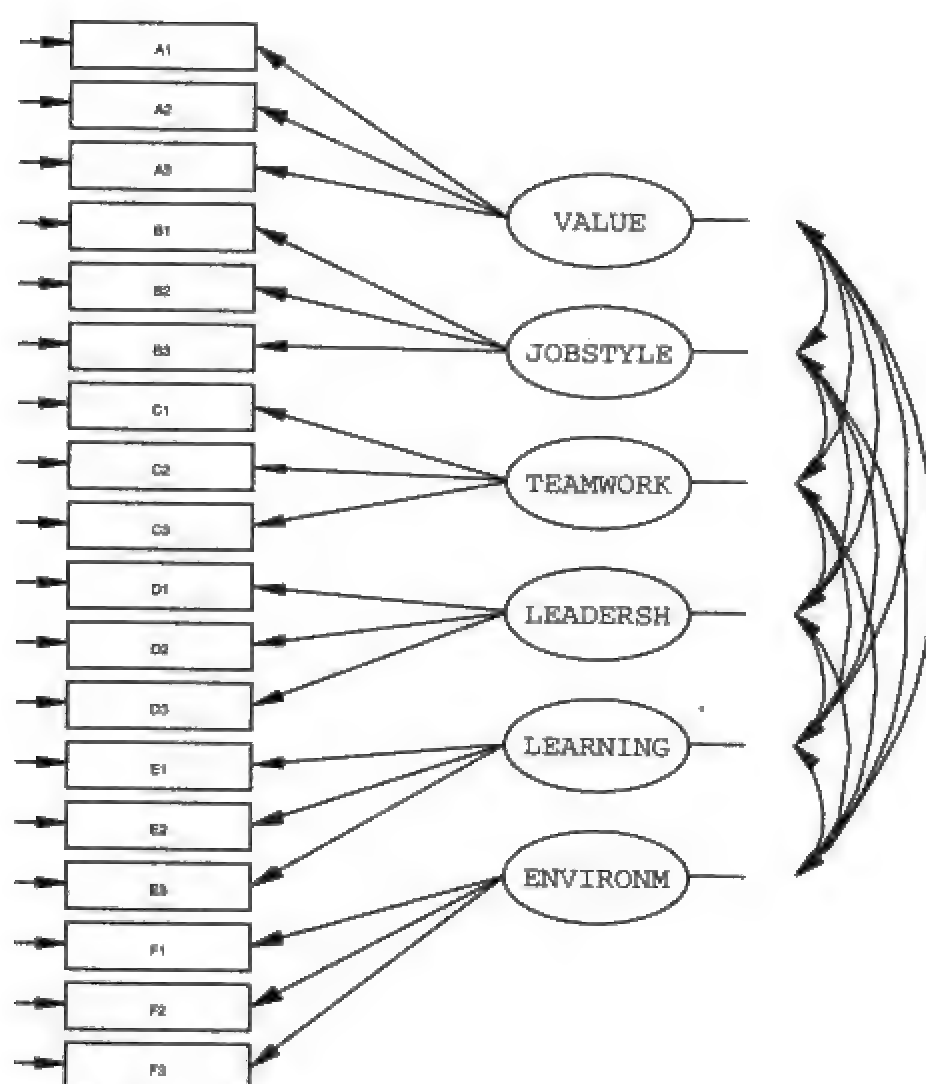


图 5.4 CFA 测量模型的假设模型

1. 模型中有 18 个测量变量 (X_1 到 X_{18} 被标签为 A1 至 F3) 与 6 个潜在变量 (ξ_1 至 ξ_6)。
2. 模型中有 18 个测量残差 (δ_1 到 δ_{18})，其变异量被自由估计。
3. 为了使 6 个潜在变量（因素）的量尺得以确立，因此每一个因素的方差被设定为 1.00。
4. 每一个测量变量仅受单一潜在变量影响（单维假设），产生 18 个因素载荷参数 (λ_1 到 λ_{18})。
5. 因素共变允许自由估计，产生 15 个相关系数参数 (ϕ_{21} 到 ϕ_{65})。
6. 测量残差之间被视为独立而没有共变。

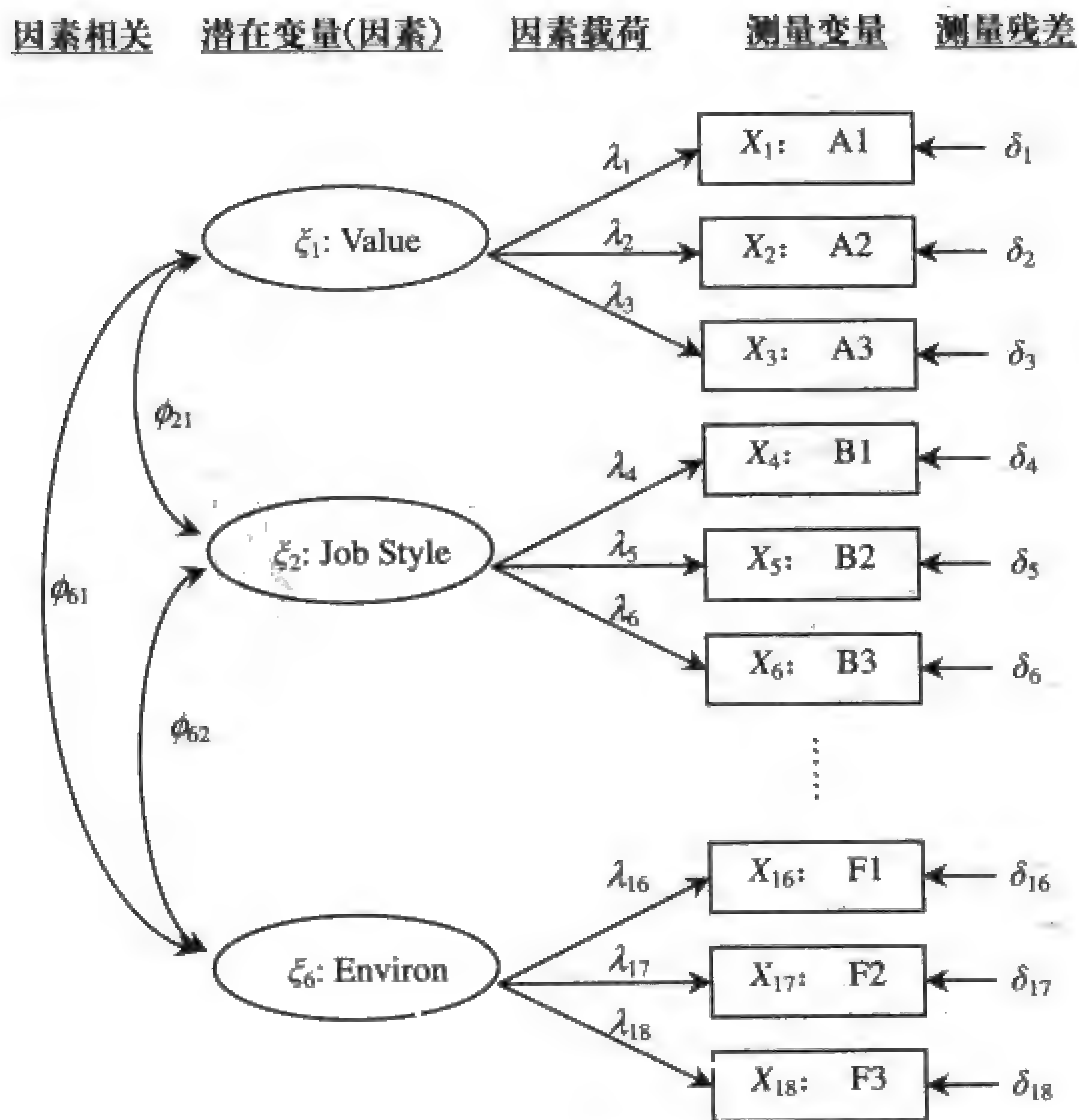


图 5.5 CFA 模型各参数路径图

(二) LISREL 数据结构

为了便于说明,本范例的分析以协方差矩阵作为输入数据,并以外挂的方式将矩阵数据存放于 ch5a.cov 档案中。在执行 LISREL 分析时,仅需指明该数据文件,并将 LISREL 语法文件 (*.ls8) 与数据文件放置于同一个目录区, LISREL 即可顺利抓取该数据文件。值得注意的是该档案的变量数目应符合 LISREL 语法中的变量数目。

Ch5a.cov 档案的内容为 18 个测量变量的方差与协方差矩阵,纵列与横行变量数目各为 18,为一对称矩阵,但因对角线上方与下方的内容相同,因此上方数据即予以省略。对角线的数字为变量自身的变异量,也就是方差。对角线下方的内容为两两变量的协方差。ch5a.cov 档案的内容如下:

```

1.005
.651 1.197
.567 .458 1.014
.284 .360 .318 1.403
.465 .556 .323 .760 1.425
.367 .406 .321 .745 .887 1.307
.313 .373 .190 .221 .414 .343 .889
.271 .297 .184 .205 .358 .267 .629 .770
.256 .329 .168 .233 .362 .288 .507 .550 .877
.326 .339 .363 .406 .554 .442 .402 .385 .406 1.009
.345 .472 .368 .373 .510 .421 .382 .370 .448 .800 1.089
.242 .418 .259 .316 .437 .374 .343 .339 .360 .613 .697 .907
.415 .339 .398 .194 .414 .316 .250 .289 .323 .322 .376 .308 1.379
.479 .412 .347 .165 .311 .286 .217 .252 .274 .308 .378 .303 .577 1.243
.471 .416 .377 .203 .351 .241 .164 .198 .238 .291 .336 .257 .503 .721 .936
.305 .278 .246 .309 .436 .405 .421 .338 .384 .447 .415 .411 .325 .219 .240 .951
.304 .302 .254 .324 .465 .386 .377 .316 .370 .407 .395 .395 .363 .233 .247 .668 .969
.366 .310 .344 .282 .548 .448 .368 .297 .340 .513 .455 .445 .494 .370 .357 .612 .755 1.118

```

测量变量的方差与协方差矩阵可以利用 LISREL 的 PRELIS 软件或其他统计软件制作, 例如 SPSS、SAS、MINITAB 等。

(三) LISREL 语法

输入数据准备完成后, 即可以进行 LISREL 语法的编写, 如表 5.3 所示。表 5.3 的 LISREL 语法, 共有 13 个主要的指令。第 1 个主指令是整个语法的标题, 以 TITLE 指令为行首。第 2 至第 4 个指令功能为数据设定, 其中第 2 个指令为 DATA, 告知变量的数目为 18 (NI=18) 与样本数目为 313 (NO=313), MA=CM 表示本范例的数据为协方差矩阵。第 3 个指令为数据内容之说明, 以 CM 为首, 表示为协方差矩阵, SY 代表为对称矩阵, 档案位置以 FI 指令说明。

表 5.3 验证性因素分析的 LISREL 语法范例 (ch5a.ls8)

1	Title ch5a Confirmatory Factor Analysis using LISREL syntax
2	DATA NI=18 NO=313 MA=CM
3	CM SY FI=ch5a.cov
4	LA; A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3
5	MODEL NX=18 NK=6 LX=FU,FI TD=DI,FR
6	LK; VALUE JOBSTYLE TEAMWORK LEADERSHIP LEARNING ENVIRONMENT
7	FREE LX(1,1) LX(2,1) LX(3,1) LX(4,2) LX(5,2) LX(6,2) LX(7,3)
8	FR LX(8,3) LX(9,3) LX(10,4) LX(11,4) LX(12,4) LX(13,5) LX(14,5)
9	FR LX(15,5) LX(16,6) LX(17,6) LX(18,6)
10	FI PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
11	VA 1 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
12	PD
13	OUTPUT SE TV RS MR FS SS SC MI

注：表中左侧的数字仅为本书作为标示行数之用，在 LISREL 语法中无该栏数字。

第 4 个指令则为这 18 个输入变量卷标，分别为 A1 至 F3，以说明这 18 个测量变量的内容。使用者可以根据自己的偏好输入适当的卷标（中文卷标亦可）。如果没有这行指令，LISREL 将自动产生 V1 至 V18 作为测量变量的卷标。

第 5 至第 11 个指令为模型界定指令。第 5 个指令（MODEL 指令）设定模型当中的外源测量变量为 18（NX=18），因素个数（外源潜在变量数目）为 6（NK=6）。测量变量被各因素解释之误差项协方差矩阵为对角矩阵且被自由估计（TD=DI,FR），亦即矩阵中只有对角在线的数据（也就是误差变异量）将被自由估计。测量变量被潜在变量解释之参数矩阵（因素载荷）为完全矩阵且被固定（LX=FU,FI）。因为大部分的因素载荷在本范例当中是假设不存在的，所以设定为 0，而被估计的因素载荷将在下面的指令 FR 指令中设定。

第 6 个主指令 LK 用于对于外源潜在变量进行卷标，六个因素分别被标签为：VALUE、JOBSTYLE、TEAMWORK、LEADERSHIP、LEARNING、ENVIRONMENT。

第 7 至第 9 个指令为详细说明各个别参数的状况。第 7 个指令 FREE（可简写为 FR）说明被估计参数为第一个测量变量（A1）与第 1 个因素（VALUE）的因素载荷 LX（1,1），第 2 个测量变量（A2）与第 1 个因素（VALUE）的因素载荷 LX（2,1），

第 3 个测量变量 (A3) 与第一个因素的因素载荷 LX (3,1), 依此类推, 直到第 18 个测量变量 (F3) 与第 6 个因素的因素载荷 LX (18,6) 被自由估计。

第 10 个指令则说明各因素当中, 被设定为 1 作为定位的参数, 分别为各外源潜在变量的方差 PH (1,1) 至 PH (6,6); 第 11 个指令以 VALUE 指令为首 (可简写为 VA), 指出先前固定的参数数值为 1, 分别为第 1 个潜在变量 (VALUE) 的方差 PH (1,1)、第 2 因素 (JOBSTYLE) 的方差 PH (2,2), 依此类推直到第 6 个因素方差 PH (6,6) 被设定为 1。

第 12 及第 13 个指令为输出指令, 第 12 个指令要求列出路径图 (Path Diagram), 第 13 个指令 OUTOUT (可简写为 OU) 为指定输出标准误 (SE)、t 检验值 (TV)、列出残差、正态化残差与 Q 图 (RS)、列出因素分素回归值 (FS)、标准化参数估计解 (SS) 与参数估计解完全标准化 (SC)、修饰指数 (MI)。如果用户不想写这么多输出指令, 可以简单的以 OUTPUT 或 OU 来要求列出 LISREL 默认的报表。

(四) SIMPLIS 语法

LISREL 的分析除了可以利用 LISREL 语法 (*.ls8) 以外, 也可以利用比较简单的 SIMPLIS 语法 (*.spl) 撰写控制指令, 其他程序则完全与 LISREL 语法相同。前述的 ex5a.ls8 改写成 SIMPLIS 语法后, 文件名为 ch5a.spl, 内容如表 5.4。

表 5.4 验证性因素分析的 SIMPLIS 语法范例 (ch5a.spl)

1	ch5a Confirmatory Factor Analysis using SIMPLIS syntax
2	Observed Variables: A1 A2 A3 B1 B2 B3 C1-C3 D1-D3 E1-E3 F1-F3
3	Covariance Matrix from File ch5a.cov
4	Sample Size = 313
5	Latent Variables VALUE JOBSTYLE TEAMWORK LEADERSH LEARNING ENVIRONM
6	Relationships
7	A1-A3 = VALUE
8	B1-B3 = JOBSTYLE
9	C1-C3 = TEAMWORK
10	D1-D3 = LEADERSH
11	E1-E3 = LEARNING

```

12  F1-F3 = ENVIRONM
13  Set the Variance of VALUE-ENVIRONM to 1
14  Path Diagram
15  LISREL Output ALL
16  End of Problem

```

表 5.4 的 SIMPLIS 语法, 共有 16 行指令。第 1 个仍为整个语法的标题, 不需以任何文字作为起始。第 2 至第 5 个指令为数据定义, 第 2 个指令以 **observed variables** 起始, 功能为指定观察变量名称, 本例观察变量为 A1 至 F3 等 18 个变量, 各变量可以逐一撰写, 也可以“-”连结有一致次序变动的相似变量, 例如 D1-D3 E1-E3 (但不可为 D1-E3, 因为英文字母变了)。第 3 个指令以 **covariance matrix** 起始, 功能在于说明数据输入的形态以及数据所在的位置。第 4 个指令以 **sample size** 起始, 说明样本数为 313。第 5 个指令以 **latent variables** 为起始, 说明潜在变量的名称。

第 6 至第 13 个指令则为模式界定, 说明本范例的检测模式, 第 6 个指令以 **relationship** 为起始, 说明观察变量与潜在变量的关系。其中第 1 至第 3 个观察变量 (A1-A3) 用来定义第一个潜在变量 (因素) (VALUE), 第 4 至第 6 个观察变量 (B1-B3) 用来定义第二个潜在变量 (JOBSTYLE), 依此类推, 直到第 12 行完成六个潜在变量的定义。

值得注意的是, 第 7 至第 12 行有多种其他写法, 例如可以每一观察变量一题, 以 18 行分别定义 18 个观察变量。例如第 7 行为 A1=VALUE, 第 8 行为 A2=VALUE, 第 9 行为 A3=VALUE, 这样的写法有一个好处是可以让每一个观察变量进行其他设定的可能 (例如如果我们要指定第一题的因素载荷为 1.0, 只要在第 7 行改为: A1=1*VALUE 即可)。另一种简要写法是把三个用来定义同一个潜在变量的观察变量写在一起, 例如: A1 A2 A3=VALUE, 结果都会一样。

第 13 个指令说明模型中被设定为 1 作为定位的参数, 本例子是以各潜在变量的方差设为 1, 因此语法为 **set the variance of VALUE to 1**, 由于潜在变量有六个, 因此以 VALUE-ENVIRONM 来一并处理其他五个, 变量名称中间以“-”作为连结。值得注意的是, 在 SIMPLIS 语法中, 此一指令可以省略, 因为 SIMPLIS 预设各因素方差为 1。

第 14 及第 15 个指令为输出指令, 第 14 个指令要求列出路径图 (Path Diagram), 第 15 个指令为指定输出形态为传统 LISREL 报表, 其中包含标准误 (SE)、t 检验值

(TV)、列出残差、正态化残差与Q图(RS)、列出因素分素回归值(FS)、标准化参数估计解(SS)与参数估计解完全标准化(SC)、修饰指数(MI)。值得注意的是,此行如果不写,则会输出简式的SIMPLIS报表。

第16个指令为结束指令,有时亦可写成end of problem。表示SIMPLIS的分析工作结束。

在SIMPLIS语法中,变量名称的大小写必须前后维持一致,其他各指令的大小写则无限制。此外,SIMPLIS指令可以近似于英文句型的方式撰写。例如样本数Sample size=313可以写为Sample size is 313或Sample size equal to 313,都可以被计算机识别。

(五) 结果报告与说明

报表(ch5a.out)的首页为系统信息与版权页,紧接着列出使用者所要求处理的LISREL指令语法。版权页与语法内容之间有一行说明,指出LISRE语法文件名与路径(from file ch5a.ls8)(注1)。

DATE: 5/18/2008
TIME: 11:30

L I S R E L 8.72

BY

Karl G. Jöreskog & Dag Sörbom

This program is published exclusively by
Scientific Software International, Inc.
7383 N. Lincoln Avenue, Suite 100
Lincolnwood, IL 60712, U.S.A.

Phone: (800)247-6113, (847)675-0720, Fax:
Copyright by Scientific Software International
Use of this program is subject to the terms
Universal Copyright Convention
Website: www.ssicentral.com

注1: LISREL 指令
档案文件名与存放
的位置。

The following lines were read from file ch5a.LS8:

Title ex5a Confirmatory Factor Analysis using LISREL syntax
DATA NI=18 NO=313 MA=CM
CM SY FI=ch5a.cov
LA; A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3
MODEL NX=18 NK=6 LX=FU,FI TD=DI,FR
LK; VALUE JOBSTYLE TEAMWORK LEADERSHIP LEARNING ENVIRONMENT
FREE LX(2,1) LX(3,1) LX(5,2) LX(6,2) LX(8,3) LX(9,3)

```
FR LX(11,4) LX(12,4) LX(14,5) LX(15,5) LX(17,6) LX(18,6)
FR LX(1,1) LX(4,2) LX(7,3) LX(10,4) LX(13,5) LX(16,6)
FI PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
VA 1 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
PD
OUTPUT SE TV RS MR FS SS SC MI
```

EX1: Confirmatory Factor Analysis for Scale development

```
Number of Input Variables 18
Number of Y - Variables 0
Number of X - Variables 18
Number of ETA - Variables 0
Number of KSI - Variables 6
Number of Observations 313
```

分析结果以使用者输入的标题作为开端，首先介绍模型当中各种变量的数目。测量变量共计 18 个，外源测量变量 (X) 数目 18，外源潜在变量 (KSI) 数目为 6，样本数为 313，其余为零。紧接着的是 LISREL 所读入的方差/协方差矩阵数据。内容如下：

Covariance Matrix						
	A1	A2	A3	B1	B2	B3
A1	0.97					
A2	0.57	1.05				
A3	0.52	0.40	0.94			
B1	0.27	0.36	0.30	1.34		
B2	0.42	0.47	0.26	0.76	1.34	
B3	0.36	0.38	0.27	0.68	0.82	1.18
C1	0.20	0.27	0.16	0.41	0.48	0.44
..... (略)						

1. 参数估计结果

接下来，LISREL 将以矩阵形态列出所有参数的估计状况，包括被自由估计以及被固定为 0 或 1 的参数。依序为 LAMBDA-X (因素载荷)、PHI (因素方差与协方差)、THETA-DELTA (测量误差方差)。其中自由估计参数依序编号由 1 至 51，表示有 51 个估计参数。

Parameter Specifications

LAMBDA-X						
	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
A1	1	0	0	0	0	0
A2	2	0	0	0	0	0
A3	3	0	0	0	0	0
B1	0	4	0	0	0	0
B2	0	5	0	0	0	0
B3	0	6	0	0	0	0
C1	0	0	7	0	0	0
C2	0	0	8	0	0	0
C3	0	0	9	0	0	0
D1	0	0	0	10	0	0
D2	0	0	0	11	0	0
D3	0	0	0	12	0	0
E1	0	0	0	0	13	0
E2	0	0	0	0	14	0
E3	0	0	0	0	15	0
F1	0	0	0	0	0	16
F2	0	0	0	0	0	17
F3	0	0	0	0	0	18
PHI						
	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
VALUE	0					
JOBSTYLE	19	0				
TEAMWORK	20	21	0			
LEADERSH	22	23	24	0		
LEARNING	25	26	27	28	0	
ENVIRONM	29	30	31	32	33	0
THETA-DELTA						
	A1	A2	A3	B1	B2	B3
	34	35	36	37	38	39
THETA-DELTA						
	C1	C2	C3	D1	D2	D3
	40	41	42	43	44	45
THETA-DELTA						
	E1	E2	E3	F1	F2	F3
	46	47	48	49	50	51

接续前述的参数估计状况，则是各参数估计的结果与统计显著性检验，依序为 LAMBDA-X、PHI、THETA-DELTA 的数据。由报表可知，本范例中，LISREL 采用

ML 法（最大概似法），总计进行了 8 次的迭代而完成了所有的参数估计。

Number of Iterations = 8

LISREL Estimates (Maximum Likelihood)

LAMBDA-X						
	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
A1	0.82 (0.05) 15.94	- -	- -	- -	- -	- -
A2	0.71 (0.06) 12.71	- -	- -	- -	- -	- -
A3	0.61 (0.05) 11.41	- -	- -	- -	- -	- -
B1	- -	0.79 (0.06) 12.74	- -	- -	- -	- -
B2	- -	0.96 (0.06) 16.56	- -	- -	- -	- -
..... (略)						

注 2：参数估计结果，依序为参数估计原始值 (0.82)、标准误 (.05)、*t* 值 (15.94)。

各参数估计结果，提供了原始估计量（非标准化数值）、标准误与统计显著性等三种资料。其中显著性检验是以 *t* 检验来进行，*t* 值越大表示强度越强。在样本数为 313 的情况下，*t* 值绝对值若超过 1.96 即可视为显著。

以 A1 变量为例，它所对应的因素为第一个因素 VALUE，因此该数值应为因素载荷 (factor loading)，其值为.82，标准误为.05，*t* 值为 15.94 (*t* 值为原始值除以标准误)。因为 *t* 值远大于 1.96，所以达到显著水平，表示该参数具有统计上的意义（注 2）。

PHI 参数的估计结果亦以矩阵形式表示，各潜在变量的方差被设定为 1，因此在对角线出现数据为 1（注 3）。下三角区域的系数可以被视为是标准化参数，各协方差的数值也就成为相关系数。

由资料可知，VALUE 与 JOBSTYLE 的相关系数相关系数为.54，*t* 值为 9.96，达显著水平（注 4）。其他各系数 *t* 值均达显著水平，表示这些标准化协方差（相关系数）的估计均具有统计的意义。

PHI		VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
VALUE	1.00						
JOBSTYLE	0.54 (0.05) 9.96	1.00					
TEAMWORK	0.49 (0.06) 8.09	0.70 (0.05) 14.87	1.00				
LEADERSH	0.42 (0.06) 7.16	0.45 (0.06) 8.12	0.52 (0.06) 9.47	1.00			
LEARNING	0.53 (0.05) 10.06	0.57 (0.05) 12.07	0.60 (0.05) 12.11	0.56 (0.05) 12.04	1.00		
ENVIRONM	0.70 (0.05) 15.13	0.39 (0.06) 6.39	0.60 (0.05) 11.00	0.32 (0.06) 5.13	0.44 (0.06) 7.93	1.00	

注 3: 各因素方差被设定为 1。
反映在对角线。

注 4: 因素间协方差估计值的
显著性检验, 协方差为 0.54、
标准误为 .05、 t 值为 9.96, 达
显著水平。

在 Theta-Delta 矩阵结果报表, 列出了各测量变量的误差变异量估计数与统计检验值。由下列的报表可知, 18 个测量变量的测量误差分别为 .30、.55、.56、.....、.26, 以 t 检验进行检定均达显著水平, 表示这些误差变异量均是有意义的估计量。

THETA-DELTA

A1	A2	A3	B1	B2	B3
0.30 (0.05) 6.50	0.55 (0.05) 9.97	0.56 (0.05) 10.68	0.72 (0.07) 10.51	0.41 (0.06) 7.20	0.45 (0.05) 8.59

THETA-DELTA

C1	C2	C3	D1	D2	D3
0.46 (0.05) 9.29	0.52 (0.06) 9.33	0.62 (0.06) 10.16	0.22 (0.03) 7.03	0.15 (0.02) 6.21	0.40 (0.04) 10.78

THETA-DELTA

E1	E2	E3	F1	F2	F3
0.29 (0.03) 9.23	0.18 (0.03) 6.09	0.33 (0.03) 9.73	0.95 (0.08) 11.38	0.51 (0.06) 8.69	0.26 (0.04) 6.25

由于先前在模型界定时,各潜在变量(各因素)的方差被设定为 1 以作为定位值,因此,此处的误差变异量也可以说是对应于潜在因素的变异量设定为 1 的标准化估计值。因此,误差变异量数值越大,表示测量残差越大,该变量被潜在变量解释的程度越低;数值愈小,表示测量残差越小,该变量被潜在变量解释的程度越高,在该因素中是重要的变量。

最后,完成各项参数估计之后,LISREL 即列出各测量变量与潜在变量的多元相关平方(Squared Multiple Correlations for X - Variables)。此一数值反映了测量变量能够被潜在变量解释的百分比,类似回归分析的 R^2 。这些数值也反映了测量变量的信度(reliability)。

Squared Multiple Correlations for X - Variables

A1	A2	A3	B1	B2	B3
0.69	0.48	0.40	0.47	0.69	0.62

Squared Multiple Correlations for X - Variables

C1	C2	C3	D1	D2	D3
0.51	0.51	0.44	0.75	0.78	0.52

Squared Multiple Correlations for X - Variables

E1	E2	E3	F1	F2	F3
0.69	0.82	0.66	0.30	0.57	0.70

以 A1 为例,有 69%的变异量能够被潜在变量解释,31%的变异量为误差变异。在各测量变量当中,以 F1 变量被解释的比率最低(30%),表示该测量变量与潜在因素的关系较微弱。在前面的残差估计中,F1 变量的误差变异量也是最高的一个变量。根据此一原理,我们可以用来诊断特定的测量变量对于因素的关联强度。

2. 模式拟合度分析

前面的分析数据为参数估计的结果,可以用来诊断个别参数的统计意义。而这些参数估计的整体效果应反映了此一模型的统计意义,可以透过各项模型拟合指数来评估,数据如下:

Goodness of Fit Statistics

Degrees of Freedom = 120
 Minimum Fit Function Chi-Square = 241.52 ($P = 0.00$)
 Normal Theory Weighted Least Squares Chi-Square = 229.33 ($P = 0.00$)
 Estimated Non-centrality Parameter (NCP) = 109.33
 90 Percent Confidence Interval for NCP = (70.37 ; 156.09)

 Minimum Fit Function Value = 0.77
 Population Discrepancy Function Value (F0) = 0.35
 90 Percent Confidence Interval for F0 = (0.23 ; 0.50)
 Root Mean Square Error of Approximation (RMSEA) = 0.054
 90 Percent Confidence Interval for RMSEA = (0.043 ; 0.065)
 P-Value for Test of Close Fit (RMSEA < 0.05) = 0.26

 Expected Cross-Validation Index (ECVI) = 1.06
 90 Percent Confidence Interval for ECVI = (0.94 ; 1.21)
 ECVI for Saturated Model = 1.10
 ECVI for Independence Model = 20.35

 Chi-Square for Independence Model with 153 Degrees of Freedom = 6312.74
 Independence AIC = 6348.74
 Model AIC = 331.33
 Saturated AIC = 342.00
 Independence CAIC = 6434.17
 Model CAIC = 573.38
 Saturated CAIC = 1153.60

 Normed Fit Index (NFI) = 0.96
 Non-Normed Fit Index (NNFI) = 0.97
 Parsimony Normed Fit Index (PNFI) = 0.75
 Comparative Fit Index (CFI) = 0.98
 Incremental Fit Index (IFI) = 0.98
 Relative Fit Index (RFI) = 0.95
 Critical N (CN) = 206.33

 Root Mean Square Residual (RMR) = 0.054
 Standardized RMR = 0.052
 Goodness of Fit Index (GFI) = 0.92
 Adjusted Goodness of Fit Index (AGFI) = 0.89
 Parsimony Goodness of Fit Index (PGFI) = 0.65

LISREL 提供了多种模型拟合指数,用以判断假设模型与实际观察值之间的拟合情形。首先,报表中列出了整体模型的自由度与卡方统计量。其中自由度为 120,卡方值为 241.52, p 值为 0.00,表示假设模型与观察值之间有显著的差异。

自由度的计算为模型的整体估计点数减去参数估计数,以本范例而言,总计 18

个测量变量可以产生 $(18 \times 19) / 2 = 171$ 个测量点数，也就是 153 对共变量与 18 个方差。而假设模型当中，共针对 51 个参数进行估计，因此剩下 120 个自由度。

卡方值有两种数值：第一为最小拟合函数卡方值（Minimum Fit Function Chi-Square），此一数值也就是用于计算所有其他各模式拟合度估计法（例如 ML 法、GLS 法）的原始卡方值。LISREL 另外提供了正态化最小平方加权卡方值（WLS chi-square）(229.33)，以及非集中化参数估计值（NCP），这两项数据均为其他模式拟合指数计算的基础值。通常在文献上所出现的卡方值，即为 WLS 卡方值。一般都希望该值越小越好，如果不显著，代表测量模型与观察值相拟合。

在各种替代性指数当中，最常用的指数为平均概似平方误根系数（RMSEA），RMSEA 系数不受样本数大小与模型复杂度的影响，当模型趋近完美拟合时，RMSEA 指数接近 0。本范例的 RMSEA 指数为 .054，90% 的置信区间为 (0.043, 0.065)，涵盖了惯用的 RMSEA = .06 的标准，显示模型拟合度佳。

其他的各参数当中，NFI、NNFI、CFI 都是常用的指数。NNFI 指数改善了 NFI 受到小样本与高自由度影响的问题。CFI 指数反映了理论模型与无任何共变关系的虚无模型（独立模型）差异程度的量数。一般而言，NFI、NNFI 与 CFI 等指数需大于 .90 才可以视为具有理想的拟合度（Hu & Bentler, 1999），本范例的 NFI (.96)、NNFI (.97)、CFI (.98) 均超过 .95 的惯用值，表示模型拟合度理想。

但是，由于 RMSEA 置信区间未能落入 .05 之内，同时非集中化参数估计值（NCP）的 95% 置信区间亦未包含 0，也就是未能涵盖最合理拟合度值，因此，若采用比较严格的说法，本范例的理论模型仍不能说具有最佳的模型拟合度。换句话说，假设模型仍有修正空间。

3. 残差分析

除了检验模型拟合度之外，衡量假设模型的计量特性，可以从各测量变量所产生的测量残差来检验。LISREL 报表，残差分析也是紧接在模型拟合指数之后。残差分析的资料，先以拟合协方差矩阵（fitted covariance matrix）列出各测量变量的方差与协方差，也就是由 LISREL 所估计出来的方差与协方差。这些数据是由模型的导出数与实际观察值之间的差距即称为残差。拟合协方差矩阵如下：

Fitted Covariance Matrix						
	A1	A2	A3	B1	B2	B3
A1	0.97					
A2	0.58	1.05				
A3	0.50	0.43	0.94			
B1	0.35	0.30	0.26	1.34		
B2	0.43	0.37	0.32	0.76	1.34	
B3	0.38	0.33	0.29	0.68	0.83	1.18
C1	0.28	0.24	0.21	0.39	0.47	0.42
.....(略)						

若取上面的数据去比较原始数据（也就是作为我们数据库的输入矩阵），我们可以看到两者之间有一定的落差，列在拟合残差（fitted residuals）当中：

Fitted Residuals						
	A1	A2	A3	B1	B2	B3
A1	0.00					
A2	0.00	0.00				
A3	0.02	-0.04	0.00			
B1	-0.08	0.06	0.04	0.00		
B2	-0.01	0.10	-0.06	0.00	0.00	
B3	-0.02	0.05	-0.01	0.00	0.00	0.00
C1	-0.09	0.02	-0.05	0.02	0.02	0.02

以 A1 与 A3 的协方差为例，实际观察值为.52（参考 Covariance Matrix 中的数据），估计值为.50，残差为.02。

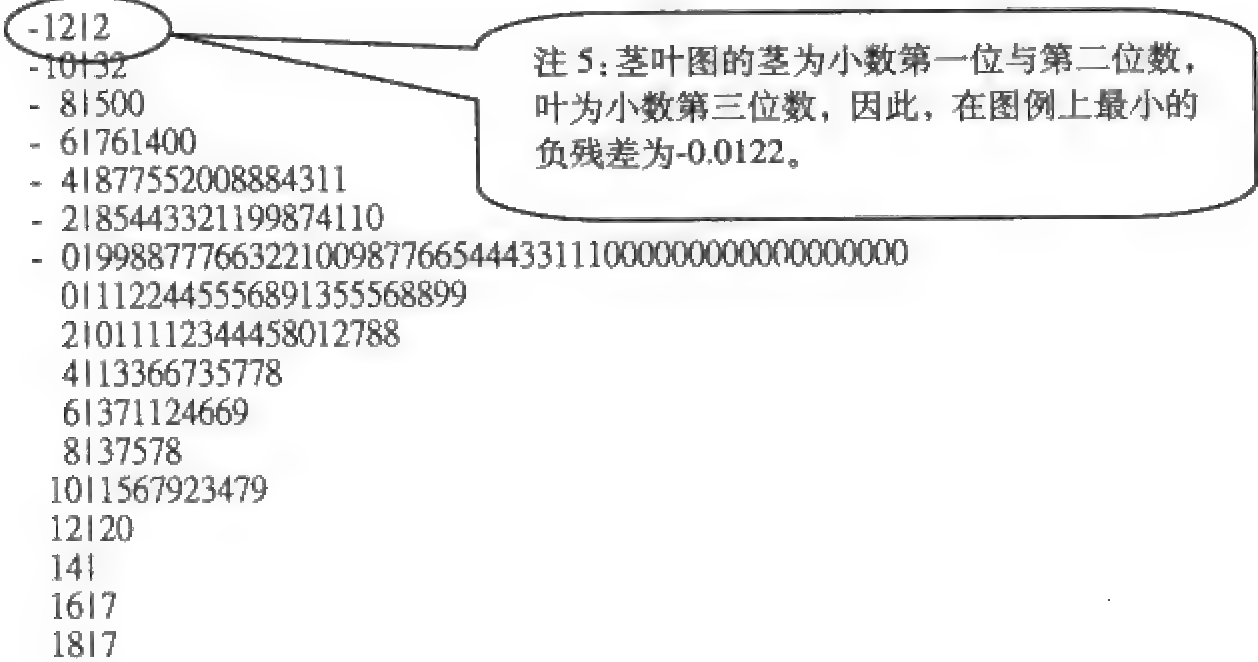
LISREL 提供的一个摘要表统整了所有 171 个观察数据点与汇出值之间的残差量，列于 Summary Statistics for Fitted Residuals 表中，同时也利用茎叶图（stemleaf plot）（注 5）绘制出残差值的分布情形。

由摘要数据可知，最小残差为-.12，最大残差为.19，残差的中位数为.00。这些资料为残差的原始估计量。

Summary Statistics for Fitted Residuals

Smallest Fitted Residual = -0.12
Median Fitted Residual = 0.00
Largest Fitted Residual = 0.19

Stemleaf Plot



Standardized Residuals

	A1	A2	A3	B1	B2	B3
A1	- .					
A2	-0.36	- .				
A3	1.89	-1.61	- .			
B1	-2.06	1.20	0.78	- .		
B2	-0.32	2.68	-1.42	0.05	- .	
B3	-0.71	1.34	-0.28	0.08	-0.14	- .
C1	-2.88	0.56	-1.41	0.55	0.52	0.79
.....(略)						

若考虑各变量的分散性以及残差的集中性与分散性, 将各残差除以 z 分数形态进行标准化, 可以得到标准化残差。最小标准化残差为-3.70, 最大标准化残差为 3.83, 标准化残差的中位数为.00。依据正态分布原理, 当标准化残差数据大于 1.96 时, 显示该残差达到.05 的统计显著水平。

Summary Statistics for Standardized Residuals

Smallest Standardized Residual = -3.70
 Median Standardized Residual = 0.00
 Largest Standardized Residual = 3.83

Stemleaf Plot

```
- 3|77
- 3|42
- 2|9555
- 2|43332110
- 1|98776665
- 1|444443332211
- 0|9998777766665555
- 0|4443333332221111000000000000000000
0|1111122244
0|5555566667777788899999
1|0223334444
1|55567889999
2|0012223444
2|567789
3|000223
3|678
```

不论是非标准化或标准化残差，可以看到有部分的残差值偏高，显示部分测量题目在假设模型的架构之下，所导出的共变量与观察值不符合，表示某特定测量题目的适切性可能有待进一步的检验。这些有待检验的测量变量可以用目测的方式自报表中找出，也可以利用 LISREL 报表的其他功能来检查。例如最大的负标准化残差-3.70，出现在 E3 与 A1 变量的共变量。报表中的条列数据列出最极端的标准化残差。

Largest Negative Standardized Residuals

Residual for	C1 and	A1	-2.88
Residual for	D2 and	C1	-3.19
Residual for	D3 and	D1	-3.66
Residual for	E2 and	D2	-3.37
Residual for	E3 and	A1	-3.70

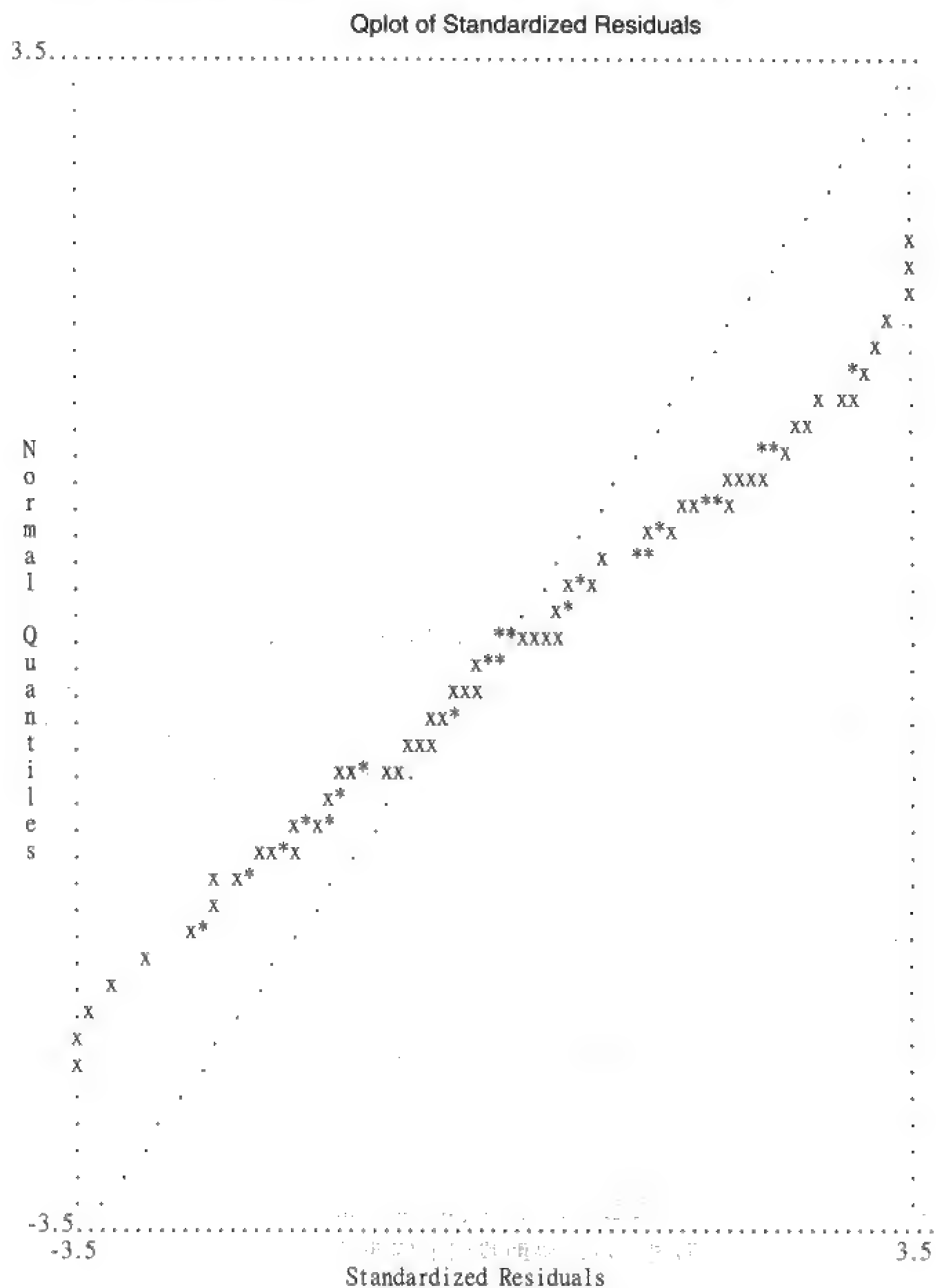
Largest Positive Standardized Residuals

Residual for	B2 and	A2	2.68
Residual for	C2 and	C1	2.66
Residual for	C3 and	A2	3.83
Residual for	D1 and	A2	3.32
Residual for	D2 and	D1	3.73

……(略)

此外，LISREL 甚至列出了残差散点图 (Qplot of Standardized Residuals)，列出标准化残差与正态化残差偏离图，可以协助我们了解残差分布的状况。分布图中的

对角线为当残差呈现正态分布时的趋势，如果实际观测到的残差过多，则在散点分布上会偏离对角线，偏离越多表示残差越非呈现正态分布。



(六) 模型修饰指数的检视

利用残差的表列与散点图, 虽然可以看到残差分布的情形, 但是无法看出测量变量与所对应的潜在变量的关系, 以及测量变量的两两共变对于模型拟合度估计的影响。尤其是那些没有被估计的参数, 可能具有统计的意义, 可以加入假设模型之中。因此, LISREL 针对非相对应的各测量变量与潜在变量的关系, 并计算出这些残差如果进行调整之后对于模型改善的贡献等统计量, 列于下列的 MI 指数中。

利用 MI 指数, 可以作为下一个步骤, 进行模型修饰的具体根据。在 LISREL 分析中, 当 MI 指数高于 5 时, 表示该残差具有修正的必要。例如 A1 变量与 LEADERSH 因素之间, MI 指数达 15.32, 显示 A1 与该因素之间, 可能具有关联, 建议纳入估计 (注 6)。也就是说 A1 变量除了被 VALUE 因素影响之外, 还可能受到 LEADERSH 因素的影响。最大的一个 MI 值是 C3 变量与 VALUE 因素之间, 若增加该因素载荷, 卡方值约可降低 14.82, 是一个很显著的改变。

Modification Indices and Expected Change

Modification Indices for LAMBDA-X

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
A1	- -	4.69	3.29	0.51	14.07	0.79
A2	- -	9.71	9.83	7.18	10.42	0.64
A3	- -	0.59	1.55	4.30	1.09	0.03
B1	1.06	- -	0.08	0.43	0.34	2.38
B2	0.44	- -	0.02	0.00	0.00	0.66
B3	0.03	- -	0.01	0.37	0.25	0.20
C1	5.22	0.56	- -	3.98	0.19	1.74
C2	1.75	1.37	- -	0.34	3.27	2.18
C3	14.82	0.20	- -	2.31	2.11	8.96
D1	3.80	1.79	0.49	- -	0.24	0.07
D2	5.86	8.64	7.01	- -	7.75	1.64
D3	0.48	4.85	19.84	- -	9.00	4.65
... ..(略)						

注 6: MI 值大于 5 时, 表示该变量的存在具有统计意义的关联性, 建议加入模型中。

除了利用 MI 指数, 进行模型修饰的具体根据。LISREL 分析提供了因素载荷预期改变 (Expected Change for LAMBDA-X)、标准化因素载荷预期改变 (Standardized Expected Change for LAMBDA-X)、完全标准化因素载荷预期改变 (Completely

Standardized Expected Change for LAMBDA-X), 来衡量每一个变量与因素关系改变的影响结果。由于本范例系以因素方差设为 1.00, 由此所获得的因素载荷其实就是标准化的数值, 因此标准化因素载荷预期改变与原始因素载荷预期改变的数值相同, 但是完全标准化 (调整各观察变量的方差后的标准化估计值) 则有所差异。

Expected Change for LAMBDA-X

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
A1	- -	-0.16	-0.14	-0.04	-0.26	0.09
A2	- -	0.22	0.23	0.16	0.21	-0.08
A3	- -	-0.05	-0.09	-0.12	0.07	-0.02
B1	-0.08	- -	-0.03	-0.04	-0.04	-0.10
B2	0.05	- -	0.02	0.00	0.00	0.05
B3	0.01	- -	0.01	0.04	0.04	0.03
C1	-0.16	0.07	- -	-0.13	0.03	-0.10
C2	0.10	-0.12	- -	0.04	-0.14	-0.12
C3	0.29	0.04	- -	0.11	0.12	0.24
D1	-0.09	0.06	-0.04	- -	0.02	-0.01
D2	-0.10	-0.12	-0.12	- -	-0.13	-0.05
D3	0.03	0.11	0.24	- -	0.16	0.10
E1	0.04	0.14	0.00	0.12	- -	-0.04
E2	0.02	-0.12	-0.03	-0.11	- -	0.04
E3	-0.07	-0.01	0.03	0.00	- -	-0.01
F1	0.00	0.10	0.28	0.23	0.18	- -
F2	-0.21	-0.12	-0.08	-0.01	-0.07	- -
F3	0.19	0.05	-0.09	-0.10	-0.04	- -

由上列数据可知, C3 与 VALUE 因素之间的关系如果纳入模型当中, 会改变因素载荷估计数达.29, 是变动最大的一个参数。有些参数的改变是减低数值, 因此数据的符号是负值。例如 F2 与 Value 的载荷数值 (-.21)。这些数值为非标准化估计值, 因此反映的是参数估计的原始数值, 参考价值较为有限, 因此我们应以标准化数值 (尤其是完全标准化数值) 来判断, 例如原始变动最大的 C3 与 VALUE 的因素载荷估计数 (.29), 标准化后仍是.28, 但是完全标准化后是.28。如果模型中新增 C3 来定义 VALUE, 所得到的完全标准化因素载荷就是.28, 减低的卡方值约为 14.82。

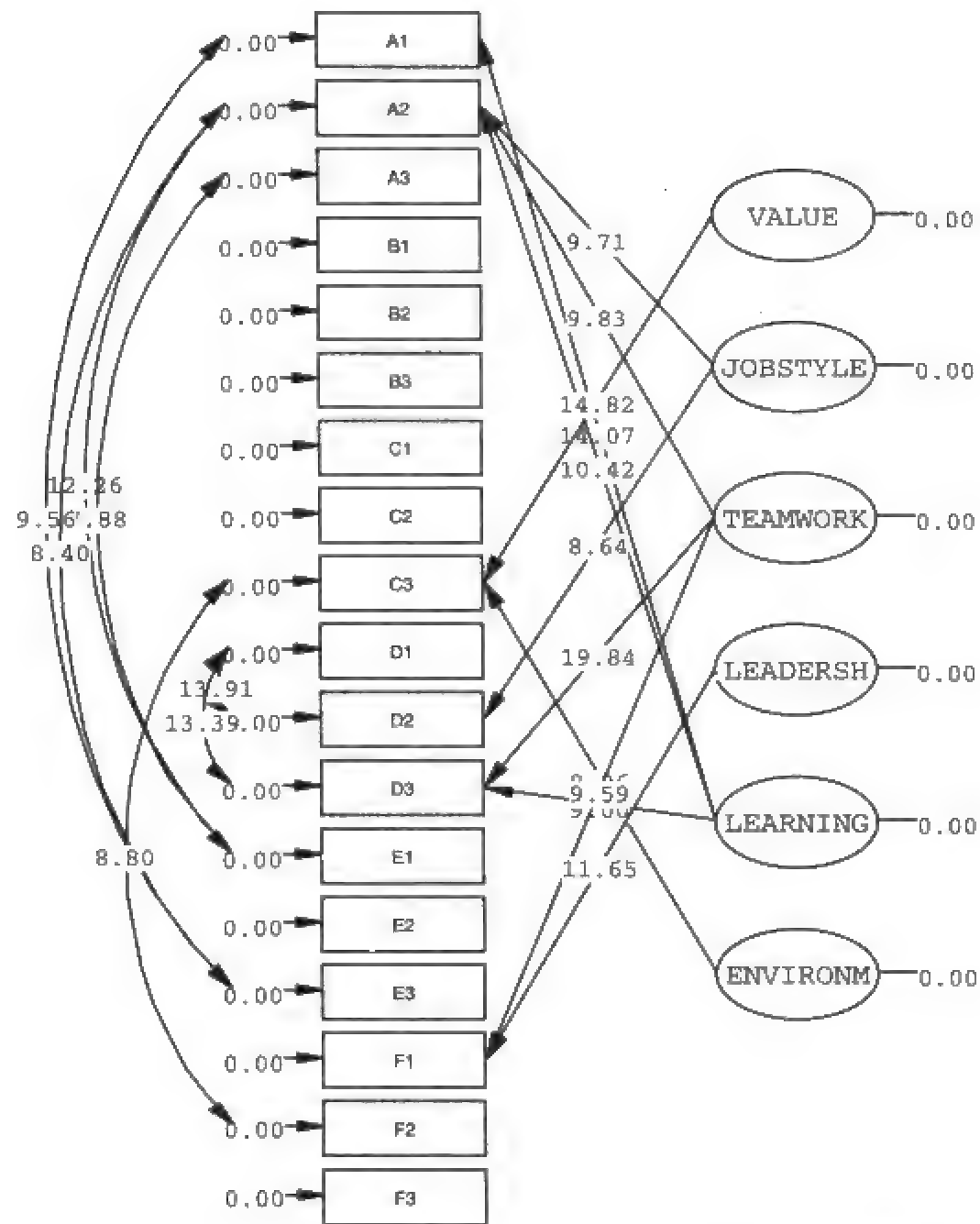
值得注意的是这些数据是单变量的统计估计值, 也就是说, 每一个参数增减的贡献是不考虑其他参数增减的影响。由此可见, 一旦增减某一个参数之后, 对其他参数的影响可能产生变动, 因此必须重新加以估计之。在 LISREL 软件, 修饰指数可以利用路径图来表示 (如图 5.6 所示), 使得模型修饰的判断更加便利。

Standardized Expected Change for LAMBDA-X

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
A1	- -	-0.16	-0.14	-0.04	-0.26	0.09
A2	- -	0.22	0.23	0.16	0.21	-0.08
A3	- -	-0.05	-0.09	-0.12	0.07	-0.02
B1	-0.08	- -	-0.03	-0.04	-0.04	-0.10
B2	0.05	- -	0.02	0.00	0.00	0.05
B3	0.01	- -	0.01	0.04	0.04	0.03
C1	-0.16	0.07	- -	-0.13	0.03	-0.10
C2	-0.10	-0.12	- -	0.04	-0.14	-0.12
C3	0.29	0.04	- -	0.11	0.12	0.24
D1	-0.09	0.06	-0.04	- -	0.02	-0.01
D2	-0.10	-0.12	-0.12	- -	-0.13	-0.05
D3	0.03	0.11	0.24	- -	0.16	0.10
E1	0.04	0.14	0.00	0.12	- -	-0.04
E2	0.02	-0.12	-0.03	-0.11	- -	0.04
E3	-0.07	-0.01	0.03	0.00	- -	-0.01
F1	0.00	0.10	0.28	0.23	0.18	- -
F2	-0.21	-0.12	-0.08	-0.01	-0.07	- -
F3	0.19	0.05	-0.09	-0.10	-0.04	- -

Completely Standardized Expected Change for LAMBDA-X

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
A1	- -	-0.16	-0.14	-0.04	-0.26	0.09
A2	- -	0.21	0.22	0.16	0.21	-0.07
A3	- -	-0.05	-0.09	-0.12	0.07	-0.02
B1	-0.07	- -	-0.02	-0.04	-0.04	-0.09
B2	0.04	- -	0.01	0.00	0.00	0.05
B3	0.01	- -	0.01	0.03	0.03	0.02
C1	-0.17	0.07	- -	-0.14	0.03	-0.10
C2	-0.10	-0.11	- -	0.04	-0.14	-0.12
C3	0.28	0.04	- -	0.10	0.11	0.23
D1	-0.09	0.07	-0.04	- -	0.03	-0.01
D2	-0.12	-0.14	-0.15	- -	-0.15	-0.06
D3	0.04	0.12	0.27	- -	0.17	0.11
E1	0.04	0.14	0.00	0.12	- -	-0.04
E2	0.02	-0.12	-0.03	-0.10	- -	0.04
E3	-0.07	-0.01	0.03	0.00	- -	-0.01
F1	0.00	0.09	0.24	0.20	0.16	- -
F2	-0.19	-0.11	-0.07	-0.01	-0.06	- -
F3	0.20	0.05	-0.09	-0.11	-0.04	- -



Chi-Square=229.33, df=120, P-value=0.00000, RMSEA=0.054

图 5.6 验证性因素分析各参数修正估计路径图

除了各测量变量与因素之间关系的修正，另一个修正途径是检验测量残差之间的相关情形。也就是检验 Theta-EPS 矩阵数值。检验的原理与参考的数据均类似。

No Non-Zero Modification Indices for PHI

Modification Indices for THETA-DELTA

	A1	A2	A3	B1	B2	B3
A1	- -					
A2	0.07	- -				
A3	3.14	4.39	- -			
B1	1.75	0.02	3.83	- -		
B2	0.28	4.81	7.16	1.37	- -	
B3	0.13	0.61	0.43	5.87	1.95	- -
.....(略)						

Expected Change for THETA-DELTA

	A1	A2	A3	B1	B2	B3
A1	- -					
A2	0.01	- -				
A3	0.08	-0.09	- -			
B1	-0.05	0.01	0.08	- -		
B2	0.02	0.08	-0.10	-0.07	- -	
B3	-0.01	-0.03	0.02	0.14	-0.11	- -
.....(略)						

值得注意的是，修正测量误差之间的相关情形，也就是说将测量误差之间的相关视为合理并纳入假设模型中，是一个具有相当争议性的做法。在 CFA 当中，任何的修饰已经存在着与理论不相符合的质疑，更何况是测量残差之间的相关，更有违传统上对于测量误差的基本独立性假设的要求。因此，建议使用者应谨慎为之，或提出有力的主张后再进行测量残差的修饰。

(七) 分析结果与参数报告

到目前为止，LISREL 分析应用于 CFA 检验已经到了一个段落，如果研究者决定不进行任何模型的修补，即可报告最终的数据。

在假设模型当中，研究者所关心的仅是测量变量与所对应的潜在变量的关系，因此，LISREL 报表中提供了各测量变量与各因素的协方差，以厘清各题目与因素之间关系的探讨。

Covariances

X - KSI

	A1	A2	A3	B1	B2	B3
VALUE	0.82	0.71	0.61	0.43	0.52	0.46
JOBSTYLE	0.44	0.38	0.33	0.79	0.96	0.86
TEAMWORK	0.40	0.35	0.30	0.55	0.67	0.60
LEADERSH	0.34	0.29	0.26	0.35	0.43	0.38
LEARNING	0.43	0.37	0.32	0.45	0.55	0.49
ENVIRONM	0.57	0.49	0.43	0.31	0.38	0.33

X - KSI

	C1	C2	C3	D1	D2	D3
VALUE	0.35	0.36	0.34	0.34	0.31	0.27
JOBSTYLE	0.49	0.51	0.49	0.36	0.33	0.29
TEAMWORK	0.70	0.74	0.70	0.42	0.39	0.34
LEADERSH	0.37	0.39	0.36	0.81	0.74	0.66
LEARNING	0.42	0.45	0.42	0.45	0.41	0.37
ENVIRONM	0.42	0.44	0.42	0.26	0.23	0.21

X - KSI

	E1	E2	E3	F1	F2	F3
VALUE	0.42	0.48	0.42	0.45	0.58	0.55
JOBSTYLE	0.46	0.52	0.45	0.25	0.32	0.31
TEAMWORK	0.49	0.55	0.48	0.38	0.50	0.47
LEADERSH	0.45	0.51	0.44	0.20	0.26	0.25
LEARNING	0.81	0.91	0.79	0.28	0.37	0.35
ENVIRONM	0.36	0.41	0.35	0.64	0.83	0.79

上述数据中，各测量变量与相对应的因素协方差以框线表示，这些协方差在各因素方差设定为 1 的情形下成为相关系数，系数越高表示该题与该因素的关系良好，例如 B2 与第二个因素（JOB STYLE）的关系，相关达.96，如果相关系数过低，表示该题与因素关系较弱。

1. 因素载荷

一般而言，因素分析所产生的数据中，以因素载荷的报告最为重要，也就是 Lambda-X 参数估计的结果。在数学计算原理上，SEM 分析获得的 Lambda-X 参数估计与传统的因素载荷估计程序没有什么不同。但是在 CFA 当中，Lambda-X 参数估

计可以排除测量残差的影响,也可以让各测量变量与潜在因素之间具有多维的假设设定(也就是同一个测量题目受到两个因素的影响)。因此,CFA 分析结果所得到的因素载荷多较传统估计为高。

Standardized Solution

LAMBDA-X

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
A1	0.82
A2	0.71
A3	0.61
B1	..	0.79
B2	..	0.96
B3	..	0.86
C1	0.70
C2	0.74
C3	0.70
D1	0.81
D2	0.74
D3	0.66
E1	0.81	..
E2	0.91	..
E3	0.79	..
F1	0.64
F2	0.83
F3	0.79

Completely Standardized Solution

LAMBDA-X

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
A1	0.83
A2	0.69
A3	0.63
B1	..	0.68
B2	..	0.83
B3	..	0.79
C1	0.72
C2	0.71
C3	0.66
D1	0.87
D2	0.89
D3	0.72
E1	0.83	..
E2	0.91	..
E3	0.81	..
F1	0.55
F2	0.76
F3	0.84

由上述的数据可以看出，不论是标准化解或完全标准化解的大多数参数估计结果均十分良好。标准化参数估计数是令每一个潜在变量的变异都相同并等于 1。本范例由于把潜在变量的参数设为 1，事实上就已经将参数以标准化参数来处理。

完全标准化解则是除了把每一个潜在变量的变异都相同并等于 1，并要求每一个观察变量的变异也都相同并等于 1，亦即令每一个观察变量的变异情形都维持一致，以利比较（一般在传统探索性因素分析所报告的因素载荷，也是将每一个观察变量的方差调整为相同并等于 1，等同于完全标准化解的意义）。

标准化因素载荷以 JOBSTYLE 的 B2 (.96) 最高，VALUE 因素中的 A3 (.61) 最低，但仍高于 .60，完全标准化后的这两个系数估计值分别为 .83 与 .63，显示整个测量问卷的测量质量良好，各题的適切性均高。此一结果仅为模型未经修饰前的原始模型的估计结果，如果经过修饰，可能有更佳的表现。

2. 因素间相关

因素间相关的估计，在 LISREL 分析中可以自 KSI 矩阵中获得。事实上，在先前参数估计过程中，因素相关的结果已经可以从报表中得出。在最后阶段再次出现相同的报表，主要是因为本范例将因素方差设定为 1，也就是将数据设定为标准化系数（Pearson 相关），因此在估计过程中即可以得到因素相关系数。在一般情形下，在参数估计时所产生的 ETA 矩阵为未标准化参数，而相关系数必须到最后才能得到。

PHI						
	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
VALUE	1.00					
JOBSTYLE	0.54	1.00				
TEAMWORK	0.49	0.70	1.00			
LEADERSH	0.42	0.45	0.52	1.00		
LEARNING	0.53	0.57	0.60	0.56	1.00	
ENVIRONM	0.70	0.39	0.60	0.32	0.44	1.00

3. 测量残差变异量

最后产生的报表,则是各测量变量的测量残差变异量估计数,也就是 Theta-EPS 矩阵的估计结果。也如同前面因素相关的资料。

THETA-DELTA

A1	A2	A3	B1	B2	B3
0.31	0.52	0.60	0.53	0.31	0.38

THETA-DELTA

C1	C2	C3	D1	D2	D3
0.49	0.49	0.56	0.25	0.22	0.48

THETA-DELTA

E1	E2	E3	F1	F2	F3
0.31	0.18	0.34	0.70	0.43	0.30

4. 验证性因素分析的路径图

LISREL 分析的优点之一,是能够产生各种参数估计值之外,可以将各项数据以路径图的方式呈现,免除人为制图的工作。因此,使用者可以将前面的标准化解结果整理后,佐以路径图(图 5.7)来强化研究结果的说明,有效地利用办公软件协助报告的完成。

(八) 内在拟合检测

前述各步骤完成了组织创新气氛量表的验证性因素分析,确立了六因素模式的適切性。下一步即是进行测量模型内在拟合的评估。

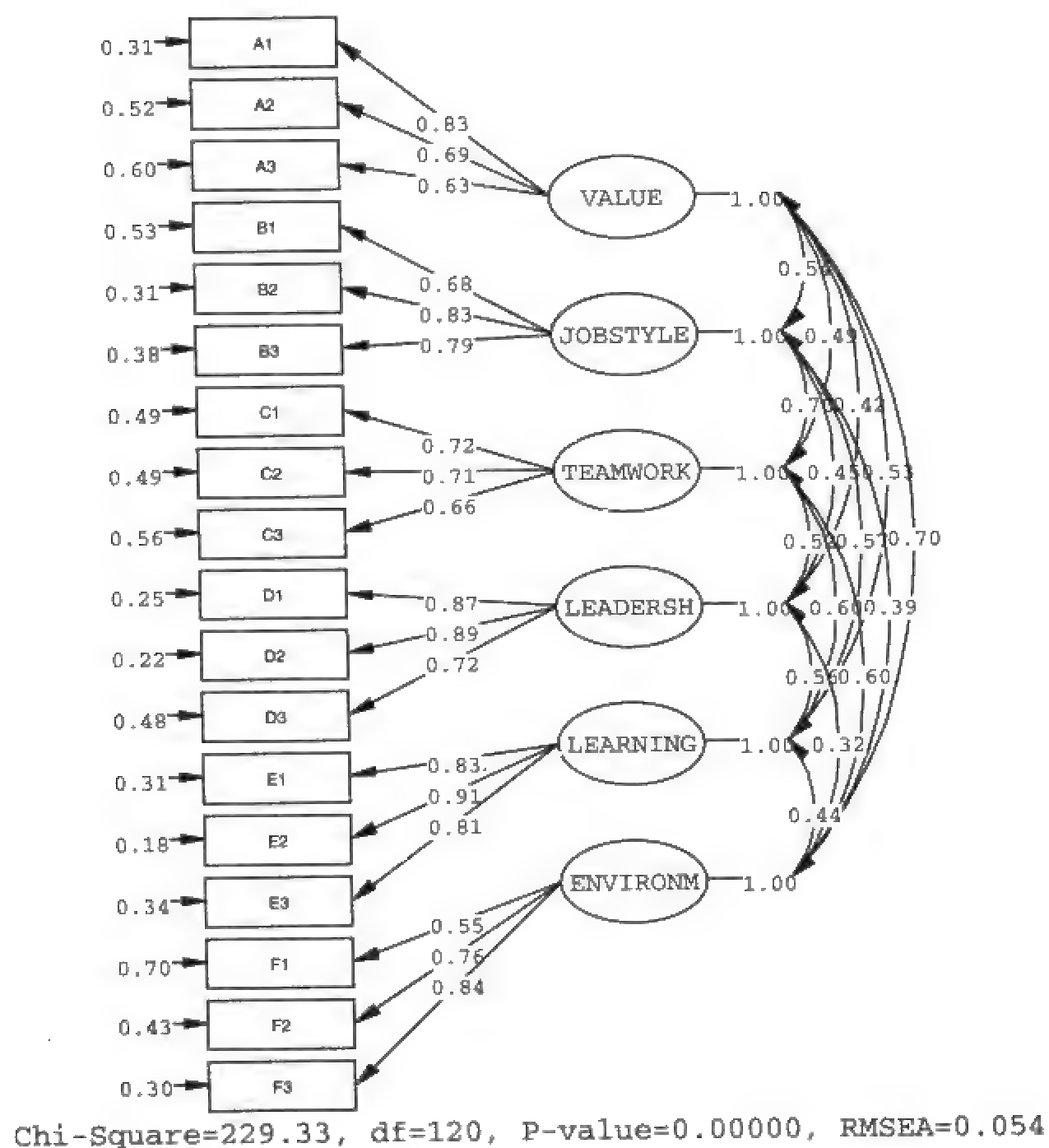


图 5.7 验证性因素分析终解路径图

表 5.5 验证性因素分析结果摘要表（修饰前模型）

因 素	题 目	λ	残差	ρ_c	ρ_v
组织 理念	1 我们公司重视人力资产、鼓励创新思考	.83	.31	.76	.52
	2 我们公司下情上达、意见交流沟通顺畅	.69	.52		
	3 我们公司能够提供诱因鼓励创新的构想	.63	.60		
工作 方式	4 当我有需要，我可以不受干扰地独立工作	.68	.53	.81	.59
	5 我的工作内容有我可以自由发挥与挥洒的空间	.83	.31		
	6 我可以自由地设定我的工作目标与进度	.79	.38		
团队 合作	7 我的工作伙伴与团队成员具有良好的共识	.72	.49	.74	.49
	8 我的工作伙伴与团队成员能够相互支持与协助	.71	.49		
	9 我的工作伙伴与团队成员能以沟通协调来化解问题 与冲突	.66	.56		
领导 效能	10 我的主管能够尊重与支持我在工作上的创意	.87	.25	.87	.69
	11 我的主管拥有良好的沟通协调能力	.89	.22		
	12 我的主管能够信任部属、适当地授权	.72	.48		
学习 成长	13 我的公司提供充分的进修机会、鼓励参与学习活动	.83	.31	.89	.72
	14 人员的教育训练是我们公司的重要工作	.91	.18		
	15 我的公司重视信息收集与新知的获得与交流	.81	.34		
环境 气氛	16 我的工作空间气氛和谐良好、令人心情愉快	.55	.70	.76	.53
	17 我有一个舒适自由、令我感到满意的工作空间	.76	.43		
	18 我的工作环境可以使我更有创意灵感与启发	.84	.30		

注：所有系数均达.05 的统计显著性。

表 5.5 列出了 18 个题目的因素载荷（完全标准化解）、残差、 ρ_c 与 ρ_v 。表中所有的参数的显著性检验均达.05 显著水平。虽然部分因素载荷的数值未达.71 的理想水平，但除了第 16 题低于.60 以外，其他题目也都能保持在良好的水平之上，因此从个别题目来看，题目的质量良好。

至于反映各因素内在拟合指数的组合信度（ ρ_c ）与变异萃取量（ ρ_v ），以公式（5.17）与公式（5.17）计算出的结果，均能达到 ρ_c 大于.7、VE 大于.50 的一般水平，显示各因素的内在拟合良好。兹以第一个因素（组织理念）为例，列出 ρ_c 与 ρ_v 的计算过程如下：

$$\rho_c = \frac{(\sum \lambda_i)^2}{[(\sum \lambda_i)^2 + \sum \Theta_{ii}]} = \frac{(.83 + .69 + .63)^2}{(.83 + .69 + .63)^2 + (.31 + .52 + .60)} = \frac{4.62}{6.05} = .76$$

$$\rho_v = \frac{\sum \lambda_i^2}{(\sum \lambda_i^2 + \sum \Theta_{ii})} = \frac{.83^2 + .69^2 + .63^2}{(.83^2 + .69^2 + .63^2) + (.31 + .52 + .60)} = \frac{1.56}{3} = .52$$

在叙述上, ρ_c 可以解释为因素内各观察变量的一致性, 其数值接近因素内各因素载荷的平均值。以组织理念因素为例, 三个因素载荷的平均为.72, 与 ρ_c 的.76 相差不远。 ρ_v 则反映了因素的解釋力, .52 的 ρ_v 表示因素能够解释观察变量 52% 的变异, 解释方式等同于传统 EFA 的特征值。

最后, 关于因素之间的区辨力, 可以相关系数的 95% 是否涵盖 1.00 来判断。由 LISREL 报表中我们可以获得各相关系数的标准误, 利用区间估计的公式, 可以计算出各相关系数的 95% 置信区间。如果这些区间没有涵盖 1.00, 可视为构念间相关具有合理区辨力的证据之一。如表 5.6。

表5.6 各因素区辨力检验摘要表

因素		组织理念	工作方式	团队合作	领导效能	学习成长
工作 方式	$r(r^2)$.54(.29)				
	95%CI	(.44,.64)				
	aveVE	.56				
团队 合作	$r(r^2)$.49(.24)	.70(.49)			
	95%CI	(.37,.61)	(.60,.80)			
	aveVE	.51	.54			
领导 效能	$r(r^2)$.42(.18)	.45(.20)	.52(.27)		
	95%CI	(.30,.54)	(.33,.57)	(.40,.64)		
	aveVE	.61	.64	.59		
学习 成长	$r(r^2)$.53(.28)	.57(.32)	.60(.36)	.56(.31)	
	95%CI	(.43,.63)	(.47,.67)	(.50,.70)	(.46,.66)	
	aveVE	.62	.66	.61	.71	
环境 气氛	$r(r^2)$.70(.49)	.39(.15)	.60(.36)	.32(.10)	.44(.19)
	95%CI	(.60,.80)	(.27,.51)	(.50,.70)	(.20,.44)	(.32,.56)
	aveVE	.53	.56	.51	.61	.63

由表 5.6 可知,各因素之间的相关系数介于.32 至.70 之间,以最大的相关.70 (团队领导与工作方式间的相关系数)为例,其 95%的置信区间 (confidence interval, CI) 为.60 至.80,并未涵盖 1.00,可以解释成该相关系数显著不等于 1.00。计算过程如下:

$$95\%CI = .70 \pm 1.96(.05) = .60 \sim .80$$

以 Fornell 与 Larker (1981) 建议的 ρ_v 比较法检验两个潜在变量的 ρ_v 平均值是否大于两个潜在变量的相关系数的平方,结果列于表 5.6 的第三列,所有各因素两两 ρ_v 的平均值均大于相关系数的平方,亦显示各构念之间具有理想的区辨力。

第四节 验证性因素分析的模型修饰

一、模型修饰的原理

如果参数估计的结果不如理想,也就是说理论模型与观察数据的拟合度不足或是参数估计的结果不理想时,研究者可以利用不同的程序与方法去修正模型,以提高模型的拟合情形或参数估计数,称为模型修饰(model modification) (Long, 1983)。虽然验证性分析的基本精神是在验证一个研究者事先提出的假设模型的合理性,模型的修正似乎违反了理论推导的基本原则,但是从分析技术的角度来看,与其报告一组本质不良的参数,不如在适当的修正模型之后,在理想模型拟合的基础之上,提出一套较具可信度的参数数据,可能更具有学术上的意义。

检查、确认并调整一个结构方程模型界定,来改善模型拟合度的整个程序,称为界定搜寻程序(specification searches)。一般来说,在进行 SEM 分析的过程当中,研究者往往有很多机会发现所提出的 SEM 模型可能有所疏忽,可能是某些变量不应用以反映某个潜在变量,或是某些参数应该纳入估计。此时,研究者可能采取某些调整或修正的步骤,使得观察数据与假设模型趋近一致,提高模型拟合度。对于测

量模型而言，可能进行调整的参数包括：

1. 改变因素载荷 (λ) 参数的固定与自由估计状态。此时可能牵涉到 Λ_x 与 Λ_y 两个矩阵的设定。
2. 改变测量残差之间的共变的估计状态 (增减测量残差 δ 或 ε 之间的共变假设)，此时将影响 Θ_δ 与 Θ_ε 两个矩阵的设定。

在上述所有可能进行调整的参数中，有些调整较为容易解释，但是有些调整可能涉及基本假设的争议，使用者应谨慎处理不同参数的调整。例如在残差部分的共变，代表研究者可能在模型当中遗漏了某一个或某几个重要的潜在变量的设定，使得不同的测量残差之间仍有共同的影响来源。而此时遗漏的潜在变量可能是另一个具有特定理论意涵的概念，或是方法效应的影响，或是其他变量的干扰，在概念厘清上相当的困难 (Bagozzi, 1983)。

二、界定搜寻程序的争议

界定搜寻程序虽然在技术上可以有效地协助研究者得到最佳的模式拟合指数，且在执行上相当单纯，任何人对于 SEM 分析稍有概念者都可以很快学会如何操作，但是使用此一程序改善模式拟合状况却有下列的疑议。

第一，此一程序使得 SEM 分析失去了验证性的特性，而带有探索性的意味。尤其当研究者所进行的调整是纯粹基于技术的考虑而无理论为基础之时，甚至会被质疑采取 SEM 分析程序处理研究者所提出问题的适当性 (Diamantopoulos, 1994)。

第二，透过界定搜寻程序所得到的 SEM 模型可以说是一个已经不同于原先所提出模型的新模型，研究者必须重新对于新模型的形成与背后原理加以说明，而不只是报告出优越的模式拟合指数，即可成为一个具有说服性的结论 (Long, 1983)。

第三，界定搜寻程序对于模型的修正与调整的判准为何？何时应进行，何时应终止搜寻，并没有清楚的界定与共识性做法。因此，利用此一程序进行的调整往往

造成俗称的过度拟合 (overfitting model) 的情况。所谓过度拟合模型是指研究者为了获得良好的拟合而以不正确的方式将模型进行调整的结果。这些调整可能是研究者加入了一些参数, 而这些参数对于模式拟合的贡献并不显著, 其效果在后续的分析中可能会消失; 或者是加入一些参数, 造成估计标准误的扩张, 影响了模型中重要参数的估计结果 (Byrne, 1998)。

过度拟合模型的问题在于加入了一些不恰当的参数或错误地移除被估计的参数, 由于模型拟合指数良好, 与实际观察数据吻合, 因此在 SEM 分析程序中不会被拒绝, 而导致错误地被接受为理想的模型, 这也是利用模型修饰程序被质疑的关键因素所在。

第四, 参数的调整会影响模型的自由度 (复杂度), 进行影响卡方值的计算, 使得参数的调整除了参数背后理论适合性的问题之外, 另外掺杂了技术上的不确定因素。

一般来说, 如果研究者移除一个参数的估计, 将增加整个模型的自由度, 如此将造成卡方值扩增的自然倾向, 卡方扩大将造成模式拟合度降低。但是, 由于模型中有较少的参数进行估计, 对于研究者有利的是符合了精简原则 (越简单的模型越佳)。相对地, 如果研究者减少参数的设限, 增加参数的估计, 将减少模型的自由度, 造成卡方值减少的自然倾向, 自动改善模式拟合度。但是, 由于模型中有较多的参数估计, 则违反了精简原则。由此可知, 参数的增减, 会使得模型精简度 (parsimony) 与拟合度 (goodness-of-fit) 的问题发生此消彼长的作用。但是, 由于模式修饰的主要目的是在改善拟合度, 因此, 一般建议使用者先增加参数的估计, 提高模式拟合度之后, 再进行参数的删减, 以简化模型的复杂度。

除了上述的争议, 界定搜寻还涉及了其他的技术与理论议题, 因此学者多建议避免盲目使用界定搜寻程序决定模型的修饰, 而宜参考其他技术性指数进行。此外, Diamantopoulos 与 Sigauw (2000) 指出了界定搜寻可能导致的各项问题, 并提出操作的建议 (表 5.7)。

表 5.7 界定搜寻的问题与操作建议

模型修饰的可能问题	操作建议
界定搜寻不一定带来一个正确合理的模型	界定搜寻不宜使用在模型验证的初期阶段（因为模型尚未稳定）
越多界定错误，界定搜寻成功的机会越少	审慎小心地进行模型界定的工作
无法拒绝一个错误模型（第二类型错误） 概率的提高	检验研究的统计检验力是否足够（例如样本数的影响），避免过度依赖卡方统计量来进行拟合度分析
测量模型的界定搜寻较不易成功，结构模型的界定搜寻较容易进行	同时兼顾测量模型与结构模型的修正可能性
以数据为基础的无限制搜寻比以理论为基础的有限度搜寻较不易成功	避免过度依赖计量指标进行修饰的决策，应适当纳入理论的考虑
先处理测量模型不见得会有帮助而可能造成修正的混淆	针对模型当中最重要、最有意义的参数进行讨论，而非取决于参数的类型
小型样本的模型修饰易造成统计结果的较大波动	若要进行模型修饰，尽可能扩大样本规模，样本规模小于 100 时不宜进行界定搜寻
模型修饰造成模型拟合指数的不理性波动	如果样本足够大，建议将样本随机切割为二，以进行交叉检验
修正后模型往往无法获得理想的效度交叉检验结果（除非样本数够大）	如有可能，可进行双重交叉检验（double cross-validation）

修改自 Diamantopoulos & Siguaw (2000), p.106

三、模型修饰的范例

以前面的 CFA 分析结果为例，MI 指数与 Expected Change for LAMBDA-X 数据

显示, C3 与组织价值 (VALUE) 因素之间的关系如果纳入模型当中, 可以有效地改善模式拟合度 (14.82), 会改变因素载荷估计数达.29, 是贡献最大的一个参数改善。因此, 我们便以 C3 与 VALUE 参数的增加作为示范。

(一) 假设模型的修饰

模型修饰的第一个步骤是修改假设模型。由于增加了 C3 与 VALUE 因素的参数估计, 也就形同在因素结构之中增加了多维的假设。对于 C3 测量变量而言, 不仅受到第三个潜在变量 TEAMWORK 因素的影响, 也受到了第一个潜在变量 VALUE 因素的影响, 显示我们将在模型当中指定 C3 变量具有多元的变异来源。

从题目的题意与因素的概念来看, C3 的内容是“我的工作伙伴与团队成员能以沟通协调来化解问题与冲突”, 这一题原本设定为“团队工作”因素之测量题目, 但受测者在这一题得分的高低也可能受到“组织价值”这一个潜在因素的影响。换言之, 组织价值越倾向于创新者, 在 C3 这一题上越可能以沟通协调来化解冲突。因此反映在 MI 指数上, 建议本题与“组织价值”因素应有关联。从表面上来看, 此一修正并无不妥, 我们可以尝试一下将此一参数纳入模型后的影响, 修饰后的假设模型路径图如图 5.8。

在此必须提醒使用者, 由于 CFA 是由理论基础推导得出的因素分析, 因此, 为了维持验证性的精神, 模型的修饰也应提出理论或概念上的说明, 避免过度修饰, 使模型修饰具有理论上的一定合理性。

(二) 模型修饰后的模型界定

模型修饰的第二个步骤是修改模型界定的内容。由于增加了 A1 与 LEADERSH 因素的参数估计, 模型界定改动情形如下:

1. 除了 C3 变量之外, 其他每一个测量变量仅受单一潜在变量影响。
2. C3 变量除了受第三个潜在变量(因素)的影响, 同时受到第一个潜在变量(因素)的影响。

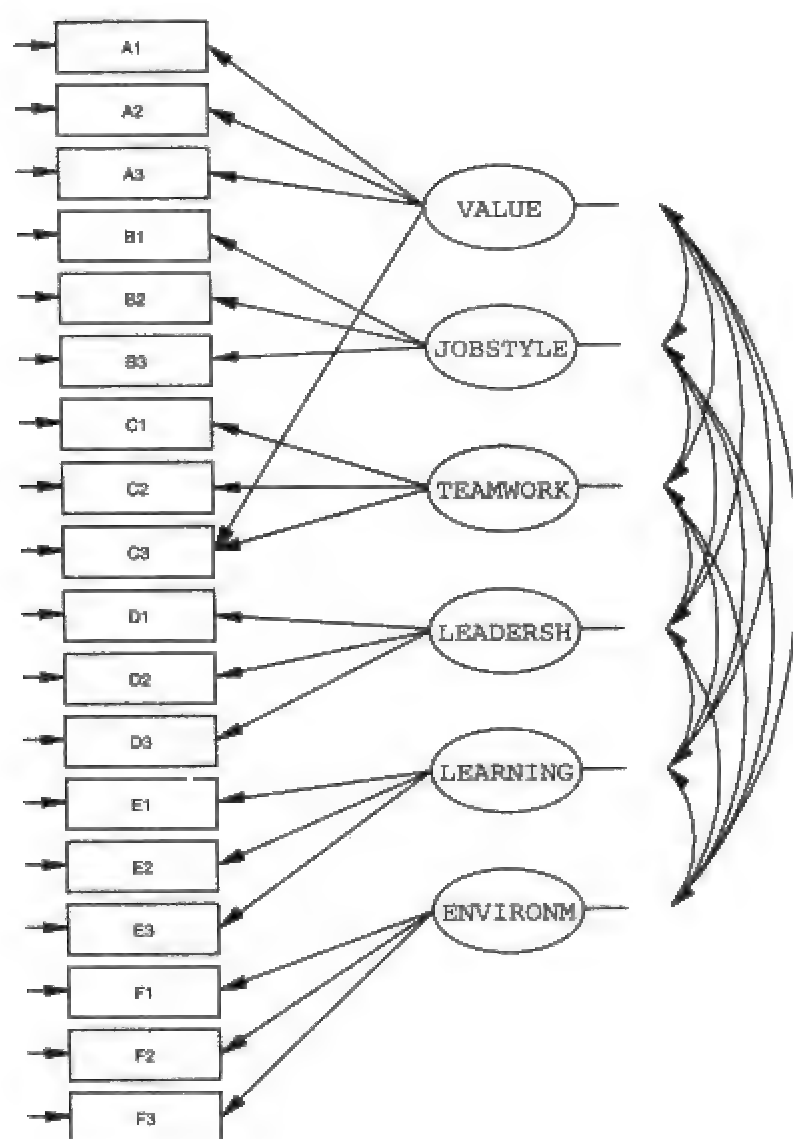


图 5.8 修饰后的假设模型图

由于模型界定条件的改变，LISREL 语法亦有了相对应的改变。在前面的路径图中，我们知道 C3 变量的改变仅是增加了一条参数的估计（自由度将减少 1），在 LISREL 语法中，也仅是增加了一个参数自由估计 `FREE LX (9,3)` 的指令于 `FREE` 指令之处，其余均无变动（修正后的 LISREL 语法文件为 `ch5a_MO.ls8`、SIMPLIS 语法文件为 `ch5a_MO.spl`）。

在 LISREL 当中，参数的变动除了可以从语法进行调整之外，也可以直接由路径图中，以鼠标点选所需进行修正的参数，拉出特性的参数路径设定各个需要增加的参数。删除的参数也可以用鼠标点选后，直接删除，操作上相当简便。

(三) 模型修饰后的结果

将上述的语法以 LISREL 软件进行分析之后,即可得到 CFA 修正后的报表(参见档案 ch5a_MO.out)。由于大多数的数据均无改变,因此下面仅就有所变动的数据进行说明。

1. 估计参数数目的变化

首先,CFA 分析结果在参数的个数上有所变动。原来的假设模型总计有 51 个参数需进行估计,修饰后的模型应有 52 个参数需进行估计,此一调整显示于 Parameter Specifications 报表页的注 7 之中。

Parameter Specifications

LAMBDA-X						
	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
A1	1	0	0	0	0	0
A2	2	0	0	0	0	0
A3	3	0	0	0	0	0
B1	0	4	0	0	0	0
B2	0	5	0	0	0	0
B3	0	6	0	0	0	0
C1	0	0	7	0	0	0
C2	0	0	8	0	0	0
C3	9	0	10	0	0	0
D1	0	0	0	11	0	0
D2	0	0	0	12	0	0
D3	0	0	0	13	0	0
E1	0	0	0	0	14	0
E2	0	0	0	0	15	0
E3	0	0	0	0	16	0
F1	0	0	0	0	0	17
F2	0	0	0	0	0	18
F3	0	0	0	0	0	19

注 7: 新增的参数估计, 为 A1 与 LEADERSH 的因素载荷, 编号为 2, 总参数数目为 52。

PHI

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
VALUE	0					
JOBSTYLE	20	0				
TEAMWORK	21	22	0			
LEADERSH	23	24	25	0		
LEARNING	26	27	28	29	0	
ENVIRONM	30	31	32	33	34	0
THETA-DELTA						
	A1	A2	A3	B1	B2	B3
	35	36	37	38	39	40
THETA-DELTA						
	C1	C2	C3	D1	D2	D3
	41	42	43	44	45	46
THETA-DELTA						
	E1	E2	E3	F1	F2	F3
	47	48	49	50	51	52

2. 新增参数估计的显著性检验

针对所加入的参数，其估计结果与显著性检验可以从 LAMBDA-X 报表页中获得。

LISREL Estimates (Maximum Likelihood)

LAMBDA-X						
	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
A1	0.81 (0.05) 15.77	- -	- -	- -	- -	- -
A2	0.71 (0.06) 12.83	- -	- -	- -	- -	- -
A3	0.62 (0.05) 11.49	- -	- -	- -	- -	- -

B1	- -	0.79 (0.06) 12.75	- -	- -	- -	- -
B2	- -	0.96 (0.06) 16.56	- -	- -	- -	- -
B3	- -	0.86 (0.06) 15.38	- -	- -	- -	- -
C1	- -	- -	0.73 (0.05) 13.58	- -	- -	- -
C2	- -	- -	0.76 (0.06) 13.26	- -	- -	- -
C3	0.27 (0.07) 4.13	- -	0.53 (0.07) 8.06	- -	- -	- -
.....(略)						

新增参数的估计结果，得到估计量为.27，标准误为.07，*t* 值为-4.13，大于.05 的显著水平门槛值 1.96，显示该参数具有统计上的意义。

至于 C3 与原因素 TEAMWORK 的因素载荷估计数，在原来的理论模型中，其值为.70，标准误为.06，*t* 值为 11.93，修饰后则降为.53，*t* 值为 8.06，显示其他的参数估计受到新增参数的影响，只有这一改变并未影响此参数为显著的这个结论，但是其他参数估计的显著性有无改变，则需加以检视。使用者必须详细重新检查所有的参数估计显著性。

3. 模型拟合度的改变

加入新的参数，主要目的在改善整体模型的拟合度，因此，大多数情况下，模型修饰的用户会直接检查模型拟合度数据的变化，以决定是否继续进行修饰。

Goodness of Fit Statistics

Degrees of Freedom = 119
Minimum Fit Function Chi-Square = 226.57 (P = 0.00)
Normal Theory Weighted Least Squares Chi-Square = 215.48 (P = 0.00)

Estimated Non-centrality Parameter (NCP) = 96.48
 90 Percent Confidence Interval for NCP = (59.26 ; 141.53)

Minimum Fit Function Value = 0.73
 Population Discrepancy Function Value (F0) = 0.31
 90 Percent Confidence Interval for F0 = (0.19 ; 0.45)
 Root Mean Square Error of Approximation (RMSEA) = 0.051
 90 Percent Confidence Interval for RMSEA = (0.040 ; 0.062)
 P-Value for Test of Close Fit (RMSEA < 0.05) = 0.43

Expected Cross-Validation Index (ECVI) = 1.02
 90 Percent Confidence Interval for ECVI = (0.90 ; 1.17)
 ECVI for Saturated Model = 1.10
 ECVI for Independence Model = 20.35

Chi-Square for Independence Model with 153 Degrees of Freedom = 6312.74
 Independence AIC = 6348.74
 Model AIC = 319.48
 Saturated AIC = 342.00
 Independence CAIC = 6434.17
 Model CAIC = 566.28
 Saturated CAIC = 1153.60

Normed Fit Index (NFI) = 0.96
 Non-Normed Fit Index (NNFI) = 0.98
 Parsimony Normed Fit Index (PNFI) = 0.75
 Comparative Fit Index (CFI) = 0.98
 Incremental Fit Index (IFI) = 0.98
 Relative Fit Index (RFI) = 0.95

Critical N (CN) = 218.30

Root Mean Square Residual (RMR) = 0.050
 Standardized RMR = 0.048
 Goodness of Fit Index (GFI) = 0.93
 Adjusted Goodness of Fit Index (AGFI) = 0.90
 Parsimony Goodness of Fit Index (PGFI) = 0.65

由数据可知，模型拟合指数有了些微的改善，参见表 5.8。修饰前，模型自由度为 120，最小拟合函数卡方值为 241.52 ($p=0.00$)，正态化最小平方加权卡方值 (WLS chi-square) 为 229.33，RMSEA 指数为.051。修饰后的模型自由度为 119，最小拟合函数卡方值为 226.57 ($p=0.00$)，WLS χ^2 为 215.48，RMSEA 指数降为.051，显示修饰后的模型拟合度更理想。卡方下降的程度与 MI 指数非常接近。

表 5.8 修正前后的模型拟合度比较表

Model	卡方检验量		df	RMSE A	NNF I	CFI	GFI	SRMR
	χ^2	WLS χ^2						
修饰前	241.52	229.33	120	.054	.97	.98	.92	.052
修饰后	226.57	215.48	119	.006	1.00	1.00	.99	.16
差异	14.95 ($p<.05$)	13.85 ($p<.05$)	1					

4. 参数估计的比较

虽然模型拟合的改善达到统计的显著水平,但参数的意义与重要性是否有所改变仍须检视参数估计的结果。由反映项目信度的相关平方 (Squared Multiple Correlations for X - Variables) 系数可以看出,修饰模型的 C3 变量的解释力并没有什么变化,修饰前为.44,修饰后为.43。从共变量来看,C3 与原因素 TEAMWORK 因素的相关为.64 (未修饰前为.70),与第一个因素 (VLAUE) 的相关为.49 (未修饰前为.34),新增参数的相关维持正相关,表示组织价值 (VALUE) 因素强度越强,C3 题得分越高。但是模型修饰后,对于原因素的相关系数强度有所减低。修饰后的报表如下:

Squared Multiple Correlations for X - Variables

C1	C2	C3	D1	D2	D3
0.56	0.54	0.43	0.75	0.78	0.52

Covariances

X - KSI

	C1	C2	C3	D1	D2	D3
VALUE	0.30	0.31	0.49	0.34	0.31	0.28
JOBSTYLE	0.49	0.51	0.50	0.36	0.33	0.29
TEAMWORK	0.73	0.76	0.64	0.40	0.36	0.32
LEADERSH	0.36	0.37	0.38	0.81	0.74	0.66
LEARNING	0.42	0.43	0.45	0.45	0.41	0.37
ENVIRONM	0.40	0.41	0.48	0.26	0.23	0.21

完全标准化解的数据则列出了新增参数的因素载荷，C3 与 VALUE 因素的载荷为.27，数值与先前 MI 指数列出的.28 很接近。C3-VALUE 参数的加入，对于原因素（TEAMWORK）当中的其他各参数因素载荷会有影响，C1、C2、C3 三个题目与 TEAMWORK 的因素载荷在修饰前为.72、.71、.66，修饰后为.75、.73、.51；而 VALUE 因素内各参数的影响则不大，修饰前 A1、A2、A3 对 VALUE 的载荷分别为.83、.69、.63，修饰后为.82、.70、.64。更简便的数据呈现方法是利用路径图，图 5.9 即列出了所有参数的标准化解（完全标准化解），其中增加的参数因素载荷为.26。

Completely Standardized Solution

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
A1	0.82	- -	- -	- -	- -	- -
A2	0.70	- -	- -	- -	- -	- -
A3	0.64	- -	- -	- -	- -	- -
B1	- -	0.68	- -	- -	- -	- -
B2	- -	0.83	- -	- -	- -	- -
B3	- -	0.79	- -	- -	- -	- -
C1	- -	- -	0.75	- -	- -	- -
C2	- -	- -	0.73	- -	- -	- -
C3	0.26	- -	0.51	- -	- -	- -
D1	- -	- -	- -	0.87	- -	- -
D2	- -	- -	- -	0.89	- -	- -
D3	- -	- -	- -	0.72	- -	- -
E1	- -	- -	- -	- -	0.83	- -
E2	- -	- -	- -	- -	0.91	- -
E3	- -	- -	- -	- -	0.81	- -
F1	- -	- -	- -	- -	- -	0.55
F2	- -	- -	- -	- -	- -	0.76
F3	- -	- -	- -	- -	- -	0.84

因素间的相关系数，在修饰前与修饰后亦有所改变，但是主要发生在新增观察变量的 TEAMWORK 因素上，与另外五个因素的相关由原来的.41、.67、.49、.57、.54 提高到.49、.70、.52、.60、.60，见表 5.9。显示在最后的标准化解中，新增观察变量影响因素的相关。

表5.9 模型修饰前后因素间相关系数比较表

	VALUE		JOBSTYLE		TEAMWORK		LEADERSH		LEARNING	
修饰前后	前	后	前	后	前	后	前	后	前	后
VALUE	1.00	1.00								
JOBSTYLE	.54	.54	1.00	1.00						
TEAMWORK	.41	.49	.67	.70	1.00	1.00				
LEADERSH	.42	.42	.45	.45	.49	.52	1.00	1.00		
LEARNING	.53	.53	.57	.57	.57	.60	.56	.56	1.00	1.00
ENVIRONM	.70	.70	.39	.39	.54	.60	.32	.32	.44	.44

整体而言,新增的参数对于整体模型的拟合度有了改善,使得 RMSEA 系数降至.051。但是新增的参数一方面增添了理论上的复杂性,使得使用者必须回到理论与文献探讨,寻求理论模型变动的解释基础;另一方面,统计资料的变动也增加了解释上的困难。

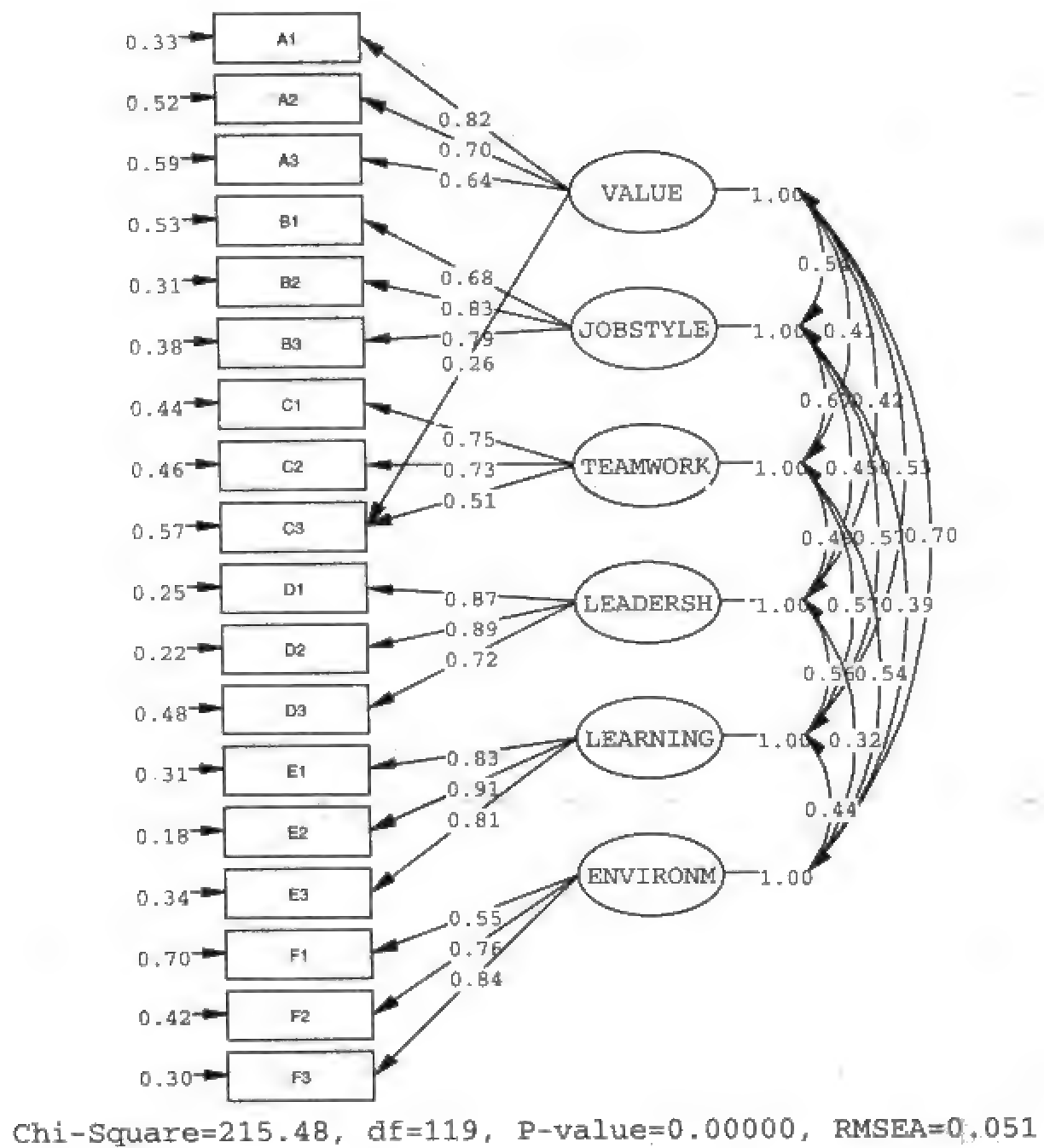


图 5.9 修饰后的验证性因素分析终解路径图

第五节 AMOS 的验证性因素分析

一、AMOS 十步骤

以 AMOS 进行 CFA 的估计分析操作方式非常简单，仅需依照下列十个步骤即可进行估计：

步骤一：绘制假设模型。依照研究者的理论架构，选择适当的图标与相关曲线，绘制完成一个 SEM 模型。

步骤二：选取数据库。打开存放观察变量的数据库。

步骤三：选取变量。依照研究者的理论模型，将各观察变量移至模型中长方形框的空白位置。

步骤四：潜在变量命名。可以逐一命名，也可以利用 AMOS 所提供的小工具（Plugins 中的 Name Unobserved Variables）快速命名，潜在变量的圈圈将自动从 F1, F2……依序命名；误差的圈圈将自动从 e1, e2……依序命名。

步骤五：点选分析的性质（Analysis properties）选择所需报表数据。例如选取模式修饰指数、标准化估计数、截距（当数据存在遗漏值时）。

步骤六：检查相关设定。例如被箭头指到的潜在变量是否有增加残差变异的设定。或是各潜在变量有无适当给定参照量尺化参数（AMOS 自动将第一条因素载荷设定为 1，如果要更改，可将鼠标移到变量或路径上，右键单击选择 Object Properties 进行更改）。

步骤七：点选计算估计（Calculate estimate）执行分析。检查模型是否收敛成功（执行之前应先开启 SRMR 对话框，使得估计完成后，同时出现 SRMR 的数据）。

步骤八：检视最终解。利用路径图标示标准化解与非标准化解（本范例组织创新气氛的 AMOS 标准化最终解路径图如图 5.10 所示）。

步骤九：检视报表。找出各项研究者所需要报告的信息。

步骤十：存档并结束分析。

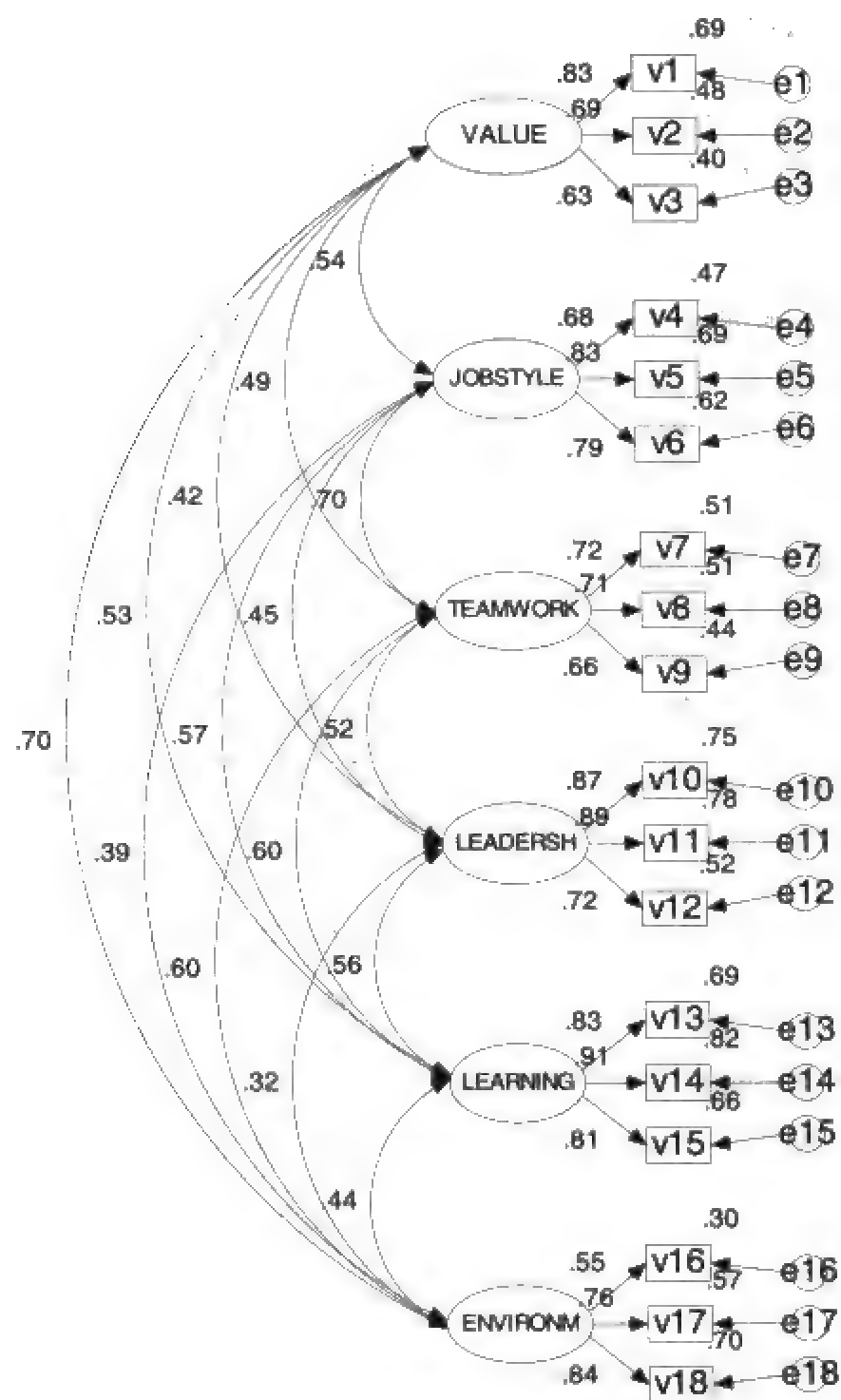
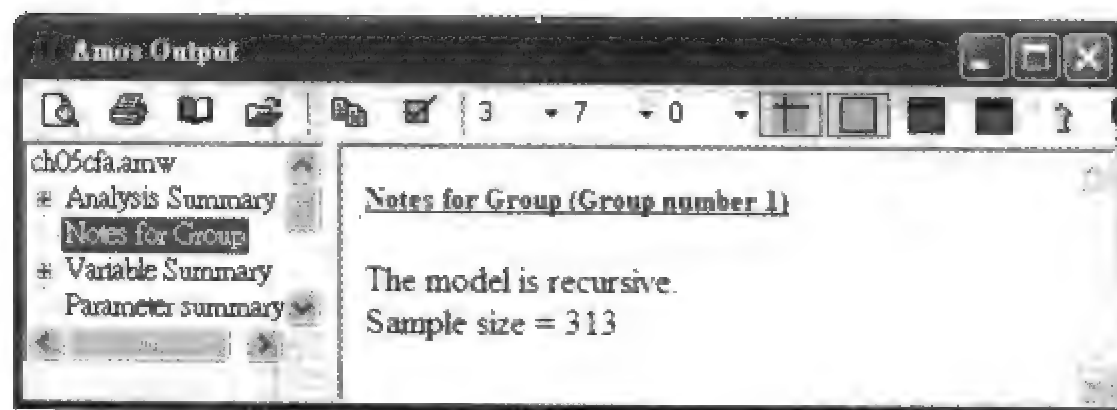
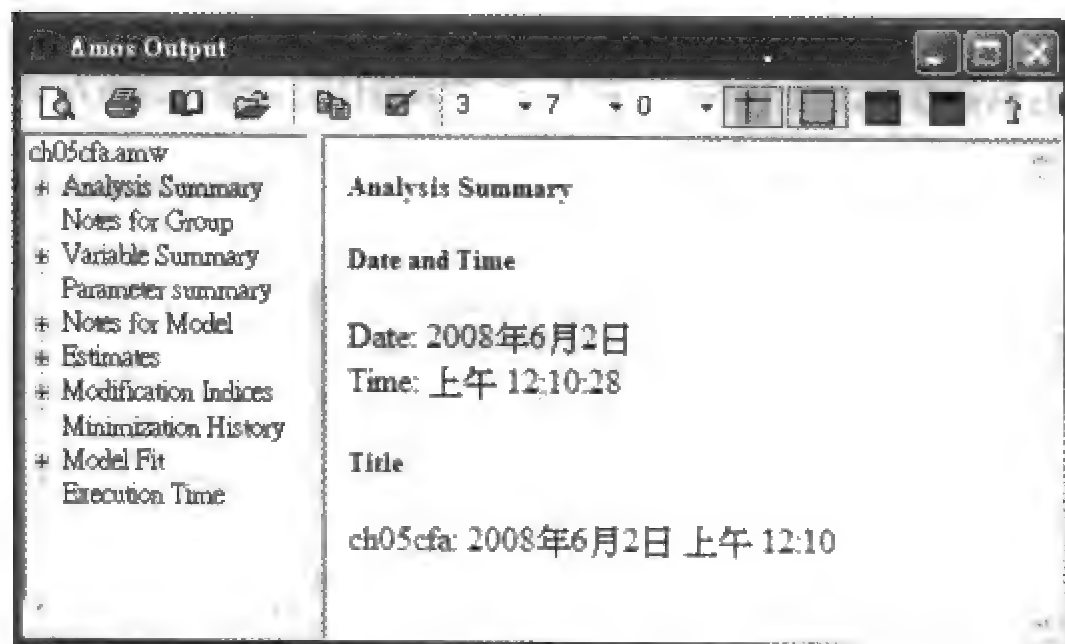


图 5.10 以 AMOS 执行 CFA 的终解路径图

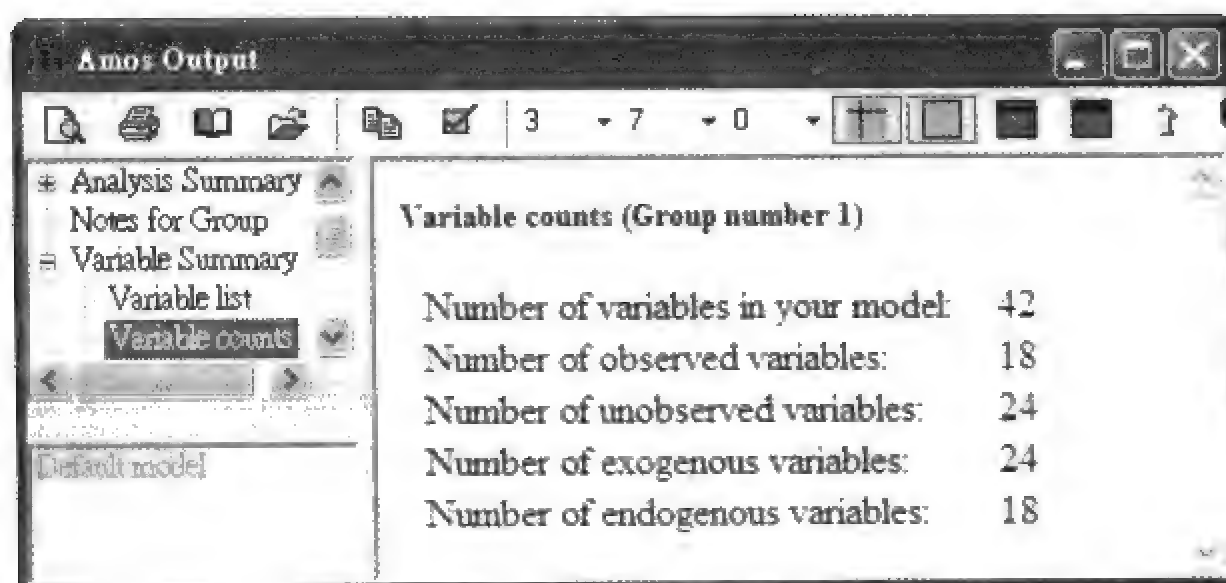
二、AMOS 的报表解读

Amos 报表有六项输出, 包含分析摘要 (Analysis Summary)、变量摘要 (Variable Summary)、模型纪录 (Notes for Model)、估计数 (Estimates)、模型拟合度 (Model Fit) 与运行时间 (Execution Time)。以下将报表内容分别列出, 并说明结果:

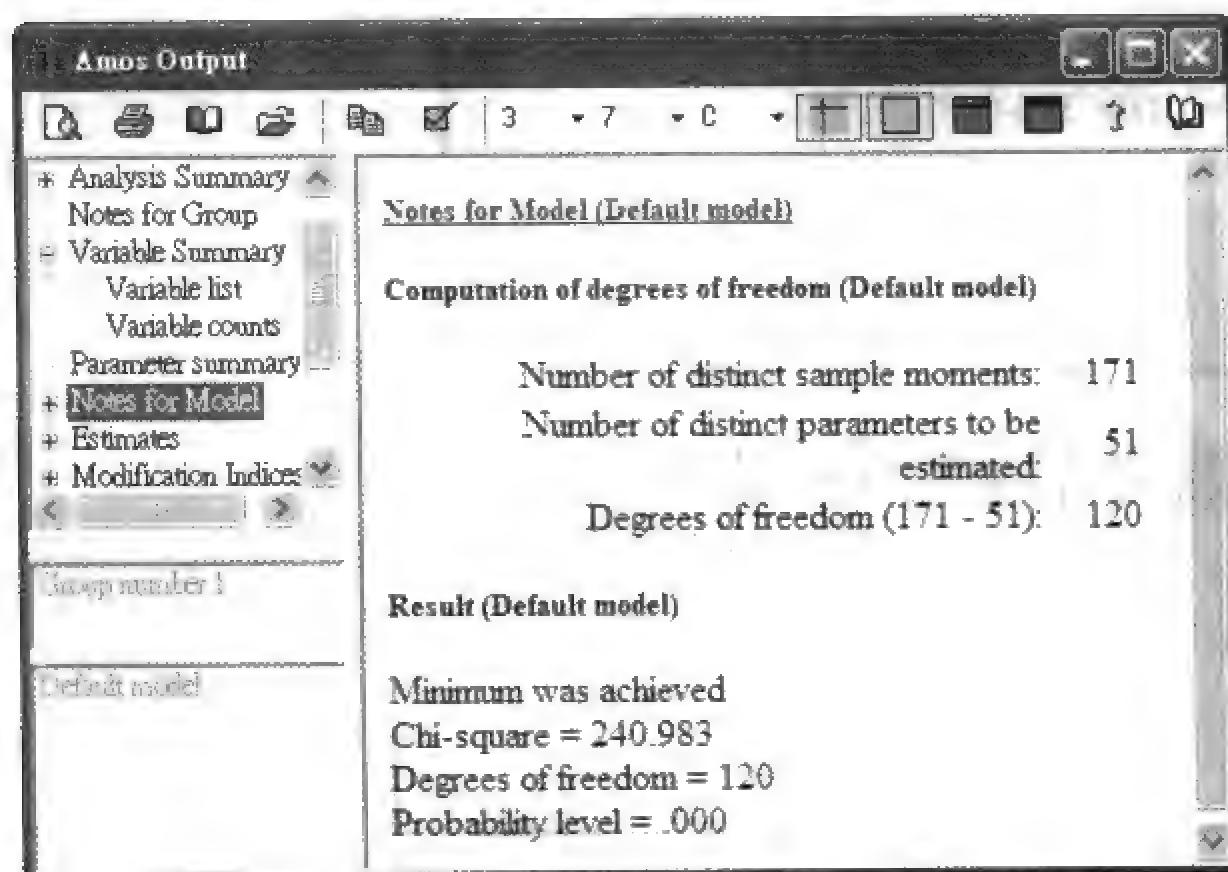
本范例已事先将遗漏值进行处理, 也就是当任何一个变量存在遗漏值, 该笔数据即完全删除; 只有完整作答的该笔数据, 才会纳入分析, 样本数共计 313 个 (以下报表档案请参阅 ch05.AmosOutput)。



变量的摘要显示模型中的所有变量, 包含观察变量 (内生变量) 与非观察变量 (外源变量), 观察变量有 18 个, 非观察变量有 24 个, 共计 42 个变量。

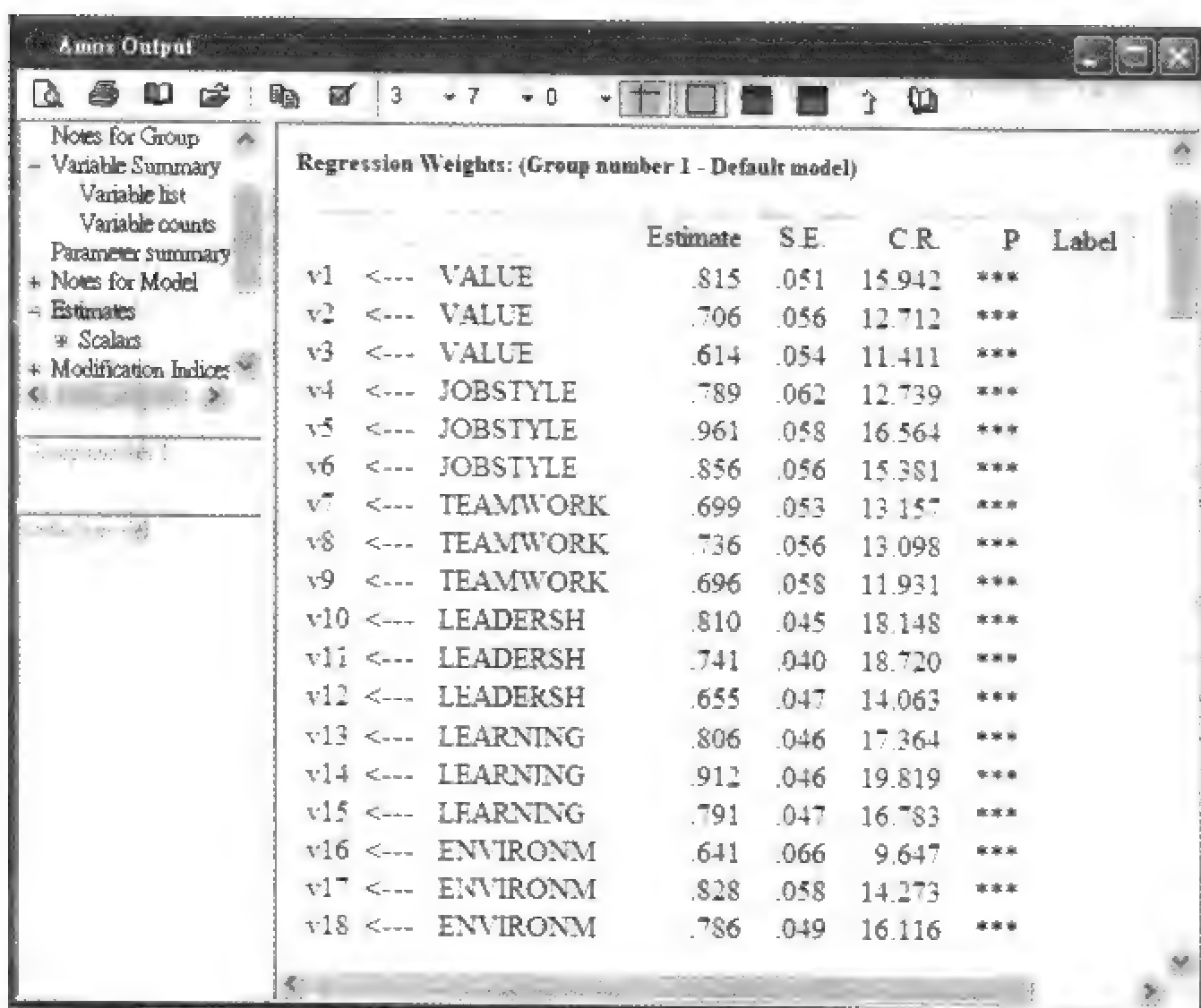


模型的纪录显示模型的测量数据数为 171 个，有 51 个参数被估计，因此自由度为 120。在模型的纪录也能看到卡方值与显著性。



Amos 的报表会列出各参数估计的结果与统计显著性检验，依序为 Regression Weights (亦即未标准化的参数估计值)、Standardized Regression Weights (亦即标准化

的参数估计值)、Covariances、Correlations、Variances 以及 Squared Multiple Correlations 的数据。本范例是采用 Maximum Likelihood Estimates (ML 法, 最大似法) 进行参数的估计。

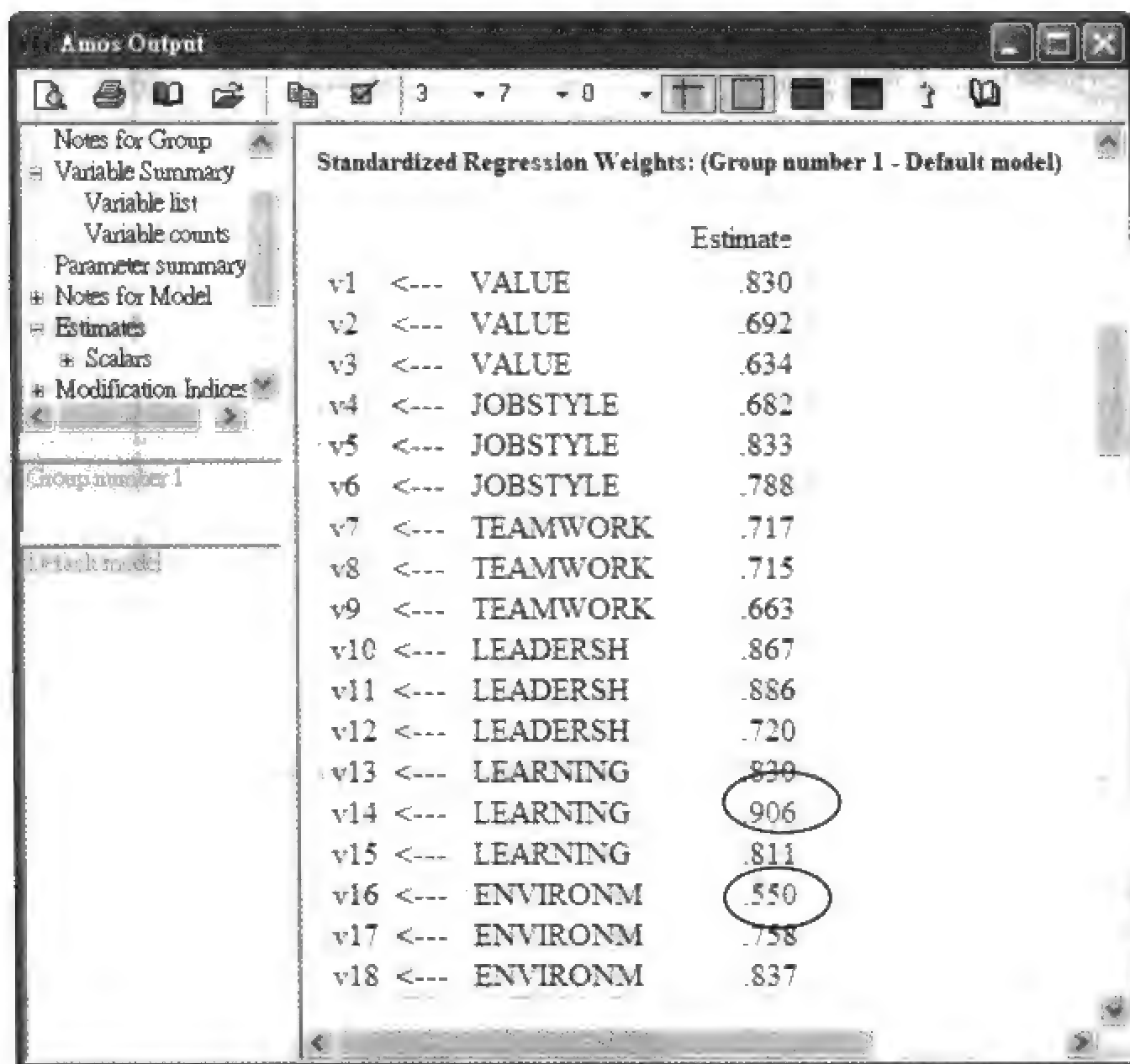


			Estimate	S.E.	C.R.	P	Label
v1	<---	VALUE	.815	.051	15.942	***	
v2	<---	VALUE	.706	.056	12.712	***	
v3	<---	VALUE	.614	.054	11.411	***	
v4	<---	JOBSTYLE	.789	.062	12.739	***	
v5	<---	JOBSTYLE	.961	.058	16.564	***	
v6	<---	JOBSTYLE	.856	.056	15.381	***	
v7	<---	TEAMWORK	.699	.053	13.157	***	
v8	<---	TEAMWORK	.736	.056	13.098	***	
v9	<---	TEAMWORK	.696	.058	11.931	***	
v10	<---	LEADERSH	.810	.045	18.148	***	
v11	<---	LEADERSH	.741	.040	18.720	***	
v12	<---	LEADERSH	.655	.047	14.063	***	
v13	<---	LEARNING	.806	.046	17.364	***	
v14	<---	LEARNING	.912	.046	19.819	***	
v15	<---	LEARNING	.791	.047	16.783	***	
v16	<---	ENVIRONM	.641	.066	9.647	***	
v17	<---	ENVIRONM	.828	.058	14.273	***	
v18	<---	ENVIRONM	.786	.049	16.116	***	

一般而言, 因素分析所产生的数据中, 以因素载荷的报告最为重要, 也就是 Lambda-X 参数估计的结果。在数学计算原理上, SEM 分析获得的 Lambda-X 参数估计与传统的因素载荷估计程序没有什么不同。但是在 CFA 当中, Lambda-X 参数估计可以排除测量残差的影响, 也可以让各测量变量与潜在因素之间具有多维的假设设定 (也就是同一个测量题目受到两个因素的影响); 因此, CFA 分析结果所得到的因素载荷多较传统估计为高。在 AMOS 的报表中, 各参数估计结果, 提供了原始估计量 (非标准化数值), 也就是报表中的 Regression Weights、标准误与统计显著性等三种资料。但作为最后报告的数据应为标准化的解, 也就是报表中的 Standardized

Regression Weights, 在标准化的参数解中却无法得知显著性, 研究者必须从原始估计量的报表中去获知显著性。

从参数估计结果的报表数据可以得知, 所有的参数值均达显著, 因素载荷以 v14 的.906 最高, v16 的.550 最低。显示所提出的组织创新气氛的理论模型良好。此一结果仅为未修饰的数据, 如果经过修饰, 可能有更好的结果。



Standardized Regression Weights: (Group number 1 - Default model)			Estimate
v1	<---	VALUE	.830
v2	<---	VALUE	.692
v3	<---	VALUE	.634
v4	<---	JOBSTYLE	.682
v5	<---	JOBSTYLE	.833
v6	<---	JOBSTYLE	.788
v7	<---	TEAMWORK	.717
v8	<---	TEAMWORK	.715
v9	<---	TEAMWORK	.663
v10	<---	LEADERSH	.867
v11	<---	LEADERSH	.886
v12	<---	LEADERSH	.720
v13	<---	LEARNING	.830
v14	<---	LEARNING	.906
v15	<---	LEARNING	.811
v16	<---	ENVIRONM	.550
v17	<---	ENVIRONM	.758
v18	<---	ENVIRONM	.837

Amos Output

Notes for Group
Variable Summary
Variable list
Variable counts
Parameter summary
Notes for Model
Estimates
Scalars
Modification Indices

Correlations: (Group number 1 - Default model)

因素间的协方差估计值。

			Estimate	S.E.	C.R.	P	Label
VALUE	<-->	JOBSTYLE	.542	.054	9.973	***	
JOBSTYLE	<-->	TEAMWORK	.697	.047	14.882	***	
VALUE	<-->	TEAMWORK	.494	.061	8.089	***	
TEAMWORK	<-->	LEADERSH	.522	.055	9.477	***	
LEADERSH	<-->	LEARNING	.557	.046	12.026	***	
LEARNING	<-->	ENVIRONM	.443	.056	7.931	***	
LEADERSH	<-->	ENVIRONM	.316	.062	5.129	***	
TEAMWORK	<-->	ENVIRONM	.600	.054	11.015	***	
JOBSTYLE	<-->	ENVIRONM	.391	.061	6.393	***	
VALUE	<-->	ENVIRONM	.695	.046	15.123	***	
TEAMWORK	<-->	LEARNING	.603	.050	12.108	***	
JOBSTYLE	<-->	LEARNING	.575	.048	12.072	***	
VALUE	<-->	LEARNING	.526	.052	10.064	***	
JOBSTYLE	<-->	LEADERSH	.447	.055	8.114	***	
VALUE	<-->	LEADERSH	.417	.058	7.164	***	

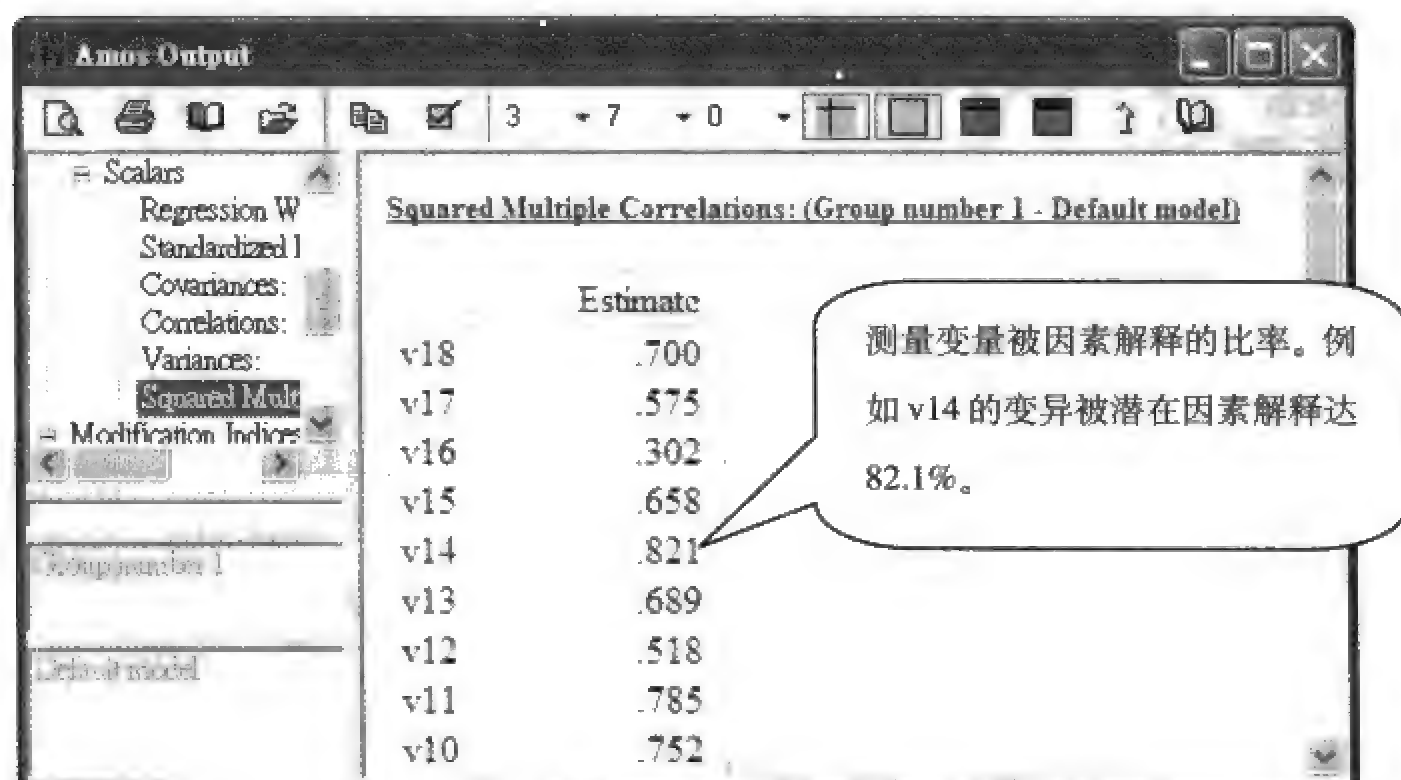
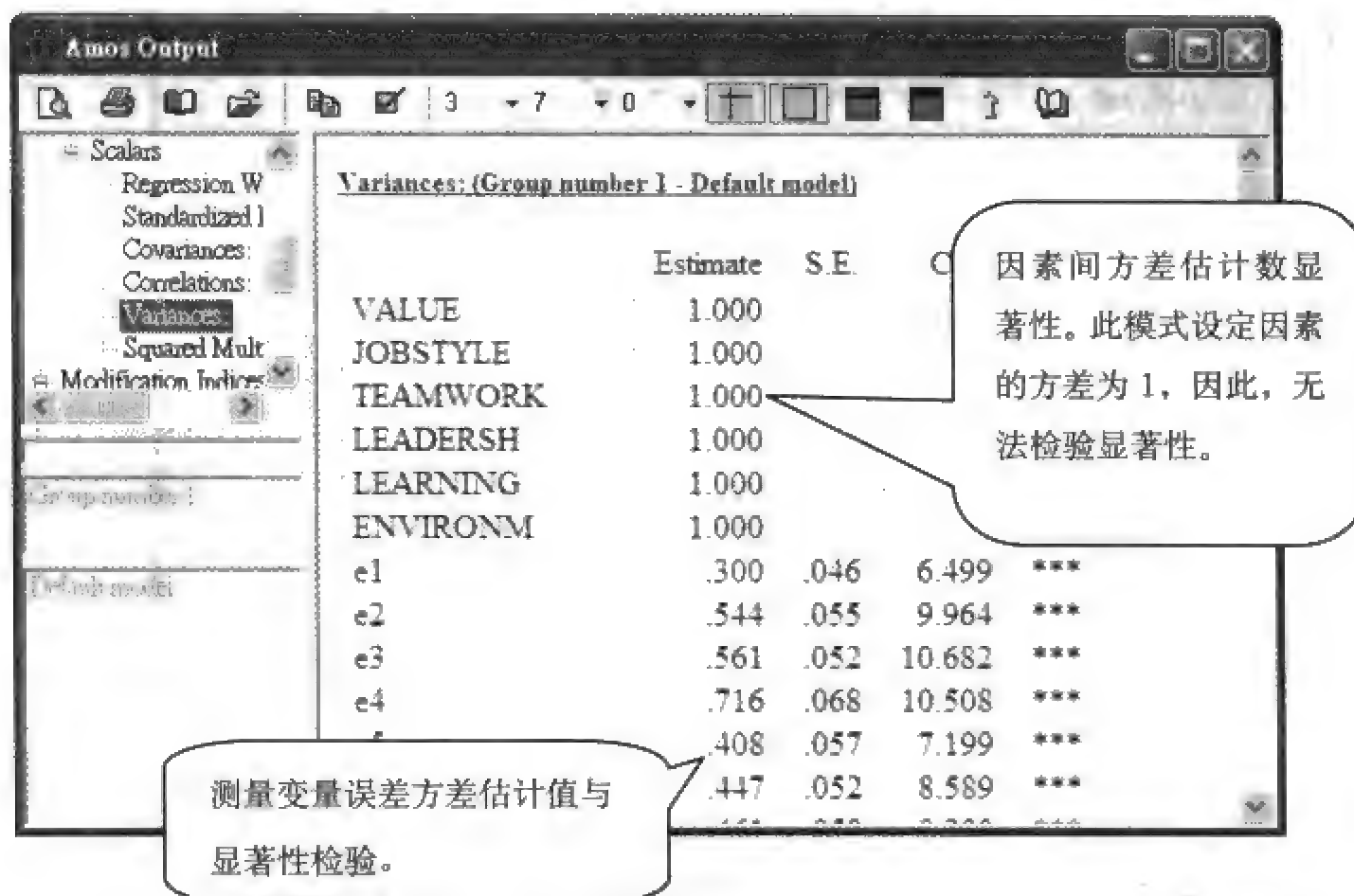
Amos Output

Notes for Group
Variable Summary
Variable list
Variable counts
Parameter summary
Notes for Model
Estimates
Scalars
Modification Indices

Correlations: (Group number 1 - Default model)

因素间的相关系数。

			Estimate
VALUE	<-->	JOBSTYLE	.542
JOBSTYLE	<-->	TEAMWORK	.697
VALUE	<-->	TEAMWORK	.494
TEAMWORK	<-->	LEADERSH	.522
LEADERSH	<-->	LEARNING	.557
LEARNING	<-->	ENVIRONM	.443
LEADERSH	<-->	ENVIRONM	.316
TEAMWORK	<-->	ENVIRONM	.600
JOBSTYLE	<-->	ENVIRONM	.391
VALUE	<-->	ENVIRONM	.695
TEAMWORK	<-->	LEARNING	.603
JOBSTYLE	<-->	LEARNING	.575
VALUE	<-->	LEARNING	.526
JOBSTYLE	<-->	LEADERSH	.447
VALUE	<-->	LEADERSH	.417



三、模型拟合度分析

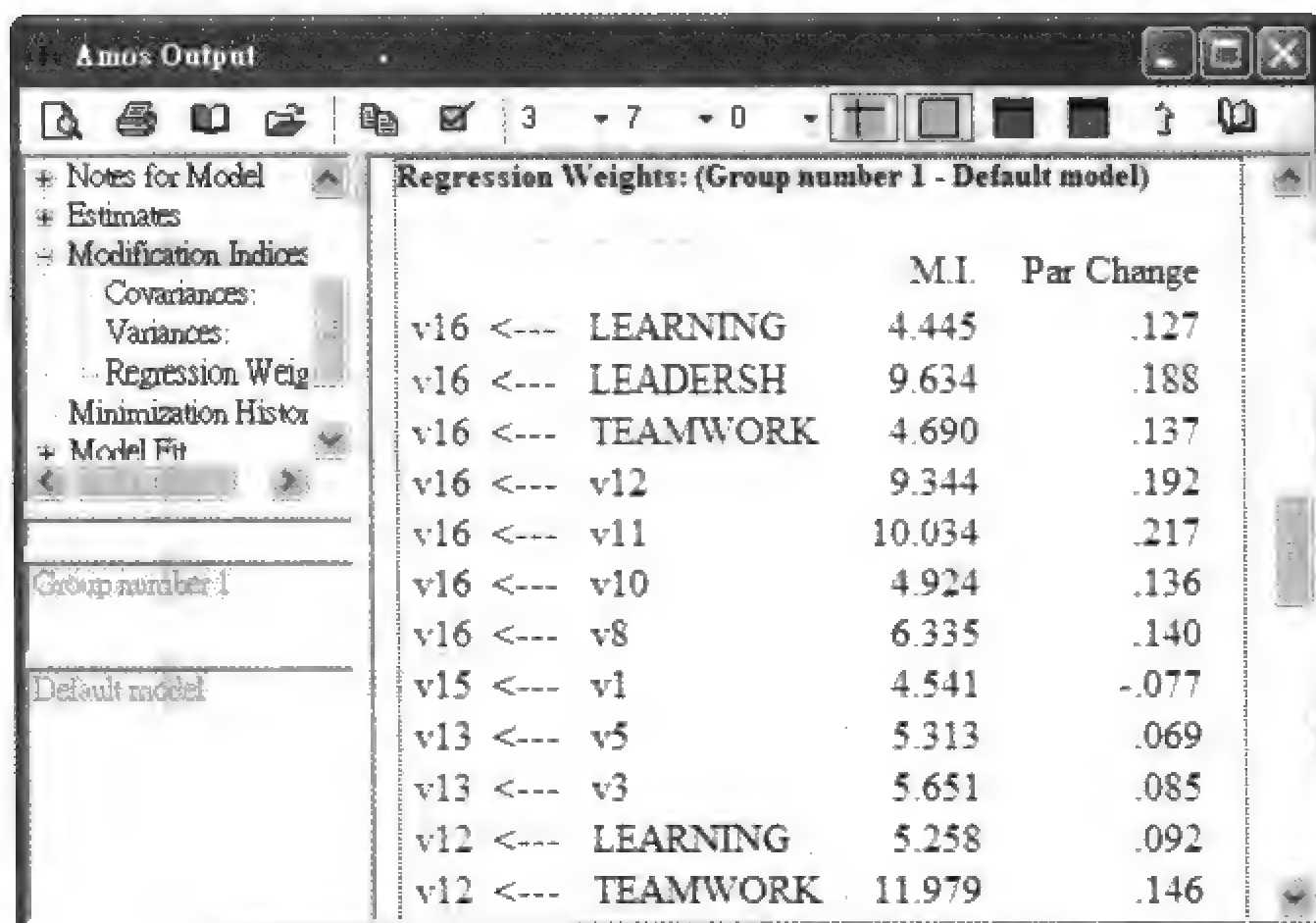
前面的参数估计透露出组织创新气氛的模型是一个不错的模型，主要的参数均达显著水平。整体效果评估也呈现相当理想的结果。

从前面的模型的纪录中显示拟合卡方值 (Chi-square) 为 240.983 ($p=.00$)，从模型拟合度摘要表中也显示 $CMIN/DF$ 为 2.008，在 AMOS 的输出报表中的 CMIN 即是指卡方值。模型拟合度报表中只呈现 RMR，未呈现标准化的 RMR，使用者必须自行至菜单栏的“Plugins”先打开“Standardized RMR”，再进行一次计算估计值 (calculate estimates)，计算机自动计算出 SRMR 的拟合指数。本范例的 $SRMR=.052$ ，平均概似平方误根系数 $RMSEA=.057$ ； TLI ($NNFI$)、 NFI 、 CFI 、 GFI 皆大于 .90，显示组织创新气氛模型具有理想的拟合度。

Model Fit Summary					
CMIN					
Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	51	240.983	120	.000	2.008
Saturated model	171	.000	0		
Independence model	18	2833.737	153	.000	18.521
RMR, GFI					
Model	RMR	GFI	AGFI	PGFI	
Default model	.053	.925	.893	.649	
Saturated model	.000	1.000			
Independence model	.353	.308	.226	.275	
Baseline Comparisons					
Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.915	.892	.955	.942	.955
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

四、模型修饰

范例所提出的模型已堪称理想,不需再进行模型修饰。如果需要进行模型修饰,可利用 MI 指数作为下一个步骤进行模型修饰的具体根据。在 AMOS 软件里,可在“分析的属性”(analysis properties)指标的 output 点选 modification index,指数内设为 4,如果选择内设值,即表示当 MI 指数高于 4 时,表示该残差具有修正的必要。例如 v12 变量与 F3 因素之间,MI 指数达 11.979,显示 v12 与该因素之间可能具有关联,建议纳入估计,也就是说 v12 变量除了被 F4 因素影响之外,还可能受到 F3 因素的影响。



Regression Weights: (Group number 1 - Default model)				
		M.I.	Par Change	
v16 <---	LEARNING	4.445	.127	
v16 <---	LEADERSH	9.634	.188	
v16 <---	TEAMWORK	4.690	.137	
v16 <---	v12	9.344	.192	
v16 <---	v11	10.034	.217	
v16 <---	v10	4.924	.136	
v16 <---	v8	6.335	.140	
v15 <---	v1	4.541	-.077	
v13 <---	v5	5.313	.069	
v13 <---	v3	5.651	.085	
v12 <---	LEARNING	5.258	.092	
v12 <---	TEAMWORK	11.979	.146	

第六节 Mplus 的验证性因素分析

Mplus 与 LISREL 都是以语法为主的分析工具。相对于 LISREL 与 AMOS 的图表功能, Mplus 的操作与呈现非常简化, 使用者只需参考指导手册的范例, 撰写适当的指令即可获得重要的数据。以下即以 Mplus(请参阅语法文件 ch5a.inp)来示范先前所说明的组织创新量表的验证性因素分析。

一、Mplus 语法

```

TITLE:      ex5a Confirmatory Factor Analysis using LISREL syntax
DATA:       FILE IS ch5a.dat;
VARIABLE:   NAMES ARE a1-a3 b1-b3 c1-c3 d1-d3 e1-e3 f1-f3;
MODEL:
  VALUE     by a1*a2-a3;
  JOBSTYLE  by b1*b2-b3;
  TEAMWORK  by c1*c2-c3;
  LEADERSH  by d1*d2-d3;
  LEARNING  by e1*e2-e3;
  ENVIRONM  by f1*f2-f3;
  VALUE@1;
  JOBSTYLE@1;
  TEAMWORK@1;
  LEADERSH@1;
  LEARNING@1;
  ENVIRONM@1;
OUTPUT:
STANDARDIZED;

```

设定模型

VALUE by a1*a2-a3 是指 VALUE 因素有三个测量指标, 其中 a1 的 λ 自由估计。若写做 VALUE by a1-a3 则 Mplus 自动设定第一条参数的 λ 为 1。

设定因素方差为 1

由于因素当中的每一条 λ 都自由估计, 因此必须设定各因素方差为 1。

打印标准化估计数

默认报表仅打印原始估计数, 若需要完全标准化解, 则需下达此一指令。

二、分析结果

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value	241.755
Degrees of Freedom	120
P-Value	0.0000

Chi-Square Test of Model Fit for the Baseline Model

Value	2842.819
Degrees of Freedom	153
P-Value	0.0000

CFI/TLI

CFI	0.955
TLI	0.942

Loglikelihood

H0 Value	-6751.785
H1 Value	-6630.907

Information Criteria

Number of Free Parameters	69
Akaike (AIC)	13641.569
Bayesian (BIC)	13900.057
Sample-Size Adjusted BIC	13681.211
$(n^* = (n + 2) / 24)$	

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.057	
90 Percent C.I.	0.047	0.067
Probability RMSEA \leq .05	0.132	

SRMR (Standardized Root Mean Square Residual)

Value	0.049
-------	-------

MODEL RESULTS

原始估计数与相关报表

t 检验

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
VALUE	BY				
	A1	0.815	0.052	15.740	0.000
	A2	0.706	0.056	12.617	0.000
	A3	0.614	0.054	11.425	0.000
JOBSTYLE	BY				
	B1	0.789	0.062	12.752	0.000
	B2	0.961	0.058	16.582	0.000
	B3	0.856	0.056	15.396	0.000

TEAMWORK BY				
C1	0.699	0.054	13.011	0.000
C2	0.736	0.057	12.965	0.000
C3	0.696	0.059	11.825	0.000
LEADERSH BY				
D1	0.810	0.044	18.277	0.000
D2	0.741	0.039	18.866	0.000
D3	0.655	0.047	14.021	0.000
LEARNING BY				
E1	0.806	0.046	17.345	0.000
E2	0.912	0.046	19.835	0.000
E3	0.791	0.047	16.845	0.000
ENVIRONM BY				
F1	0.641	0.067	9.546	0.000
F2	0.828	0.058	14.325	0.000
F3	0.786	0.049	16.094	0.000
JOBSTYLE WITH				
VALUE	0.542	0.055	9.802	0.000
TEAMWORK WITH				
VALUE	0.494	0.063	7.825	0.000
JOBSTYLE	0.697	0.047	14.824	0.000
LEADERSH WITH				
VALUE	0.417	0.059	7.068	0.000
JOBSTYLE	0.447	0.055	8.067	0.000
TEAMWORK	0.522	0.056	9.335	0.000
LEARNING WITH				
VALUE	0.526	0.054	9.677	0.000
JOBSTYLE	0.575	0.048	11.974	0.000
TEAMWORK	0.603	0.050	12.014	0.000
LEADERSH	0.557	0.047	11.851	0.000
ENVIRONM WITH				
VALUE	0.695	0.046	15.263	0.000
JOBSTYLE	0.391	0.061	6.394	0.000
TEAMWORK	0.600	0.056	10.743	0.000
LEADERSH	0.316	0.062	5.071	0.000
LEARNING	0.443	0.056	7.909	0.000
Intercepts				
A1	4.415	0.056	79.524	0.000
A2	4.307	0.058	74.595	0.000
A3	4.073	0.055	74.447	0.000
.....(略)				
Variances				
VALUE	1.000	0.000	999.000	999.000
JOBSTYLE	1.000	0.000	999.000	999.000
TEAMWORK	1.000	0.000	999.000	999.000
LEADERSH	1.000	0.000	999.000	999.000
LEARNING	1.000	0.000	999.000	999.000
ENVIRONM	1.000	0.000	999.000	999.000

方差设定为 1.0

Residual Variances				
A1	0.300	0.048	6.222	0.000
A2	0.544	0.056	9.796	0.000
A3	0.561	0.052	10.692	0.000
B1	0.716	0.068	10.512	0.000
B2	0.408	0.057	7.196	0.000
B3	0.447	0.052	8.585	0.000
.....(略)				

STANDARDIZED MODEL RESULTS

STDYX Standardization

完全标准化解与相关检验

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
VALUE	BY				
A1		0.830	0.032	26.314	0.000
A2		0.692	0.038	18.205	0.000
A3		0.634	0.041	15.518	0.000
JOBSTYLE	BY				
B1		0.682	0.036	18.754	0.000
B2		0.833	0.027	30.700	0.000
B3		0.788	0.030	26.634	0.000
TEAMWORK	BY				
C1		0.717	0.038	19.037	0.000
C2		0.715	0.038	18.935	0.000
C3		0.663	0.041	16.301	0.000
LEADERSH	BY				
D1		0.867	0.021	40.729	0.000
D2		0.886	0.020	43.694	0.000
D3		0.720	0.032	22.644	0.000
LEARNING	BY				
E1		0.830	0.023	36.901	0.000
E2		0.906	0.017	51.860	0.000
E3		0.811	0.023	34.623	0.000
ENVIRONM	BY				
F1		0.550	0.046	11.828	0.000
F2		0.758	0.033	22.837	0.000
F3		0.837	0.030	27.683	0.000
JOBSTYLE WITH VALUE		0.542	0.055	9.802	0.000
TEAMWORK WITH VALUE		0.494	0.063	7.825	0.000
JOBSTYLE		0.697	0.047	14.824	0.000
LEADERSH WITH VALUE		0.417	0.059	7.068	0.000
JOBSTYLE		0.447	0.055	8.067	0.000
TEAMWORK		0.522	0.056	9.335	0.000

LEARNING WITH				
VALUE	0.526	0.054	9.677	0.000
JOBSTYLE	0.575	0.048	11.974	0.000
TEAMWORK	0.603	0.050	12.014	0.000
LEADERSH	0.557	0.047	11.851	0.000
ENVIRONM WITH				
VALUE	0.695	0.046	15.263	0.000
JOBSTYLE	0.391	0.061	6.394	0.000
TEAMWORK	0.600	0.056	10.743	0.000
LEADERSH	0.316	0.062	5.071	0.000
LEARNING	0.443	0.056	7.909	0.000
Intercepts				
A1	4.495	0.188	23.867	0.000
A2	4.216	0.178	23.721	0.000
A3	4.208	0.177	23.716	0.000
.....(略)				
Variances				
VALUE	1.000	0.000	999.000	999.000
JOBSTYLE	1.000	0.000	999.000	999.000
TEAMWORK	1.000	0.000	999.000	999.000
LEADERSH	1.000	0.000	999.000	999.000
LEARNING	1.000	0.000	999.000	999.000
ENVIRONM	1.000	0.000	999.000	999.000
Residual Variances				
A1	0.311	0.052	5.950	0.000
A2	0.522	0.053	9.929	0.000
A3	0.598	0.052	11.554	0.000
B1	0.535	0.050	10.781	0.000
B2	0.306	0.045	6.775	0.000
B3	0.379	0.047	8.134	0.000
.....(略)				

基本上, LISREL、AMOS、Mplus 等不同软件所估计出来的参数数值几乎完全相同, 但是 AMOS 报表仅提供残差的原始估计数。

三种软件在估计模型整体拟合性时差异比较明显。LISREL 提供了两种最小拟合函数 χ^2 与 WLS χ^2 , 其中 WLS χ^2 数值较小, 也因此导致 CFI、TLI、RMSEA 等各项指数比另外两个软件更理想。更进一步的比较在此省略, 有兴趣的读者可以自行检视比较三种软件的差异。

第七节 结 语

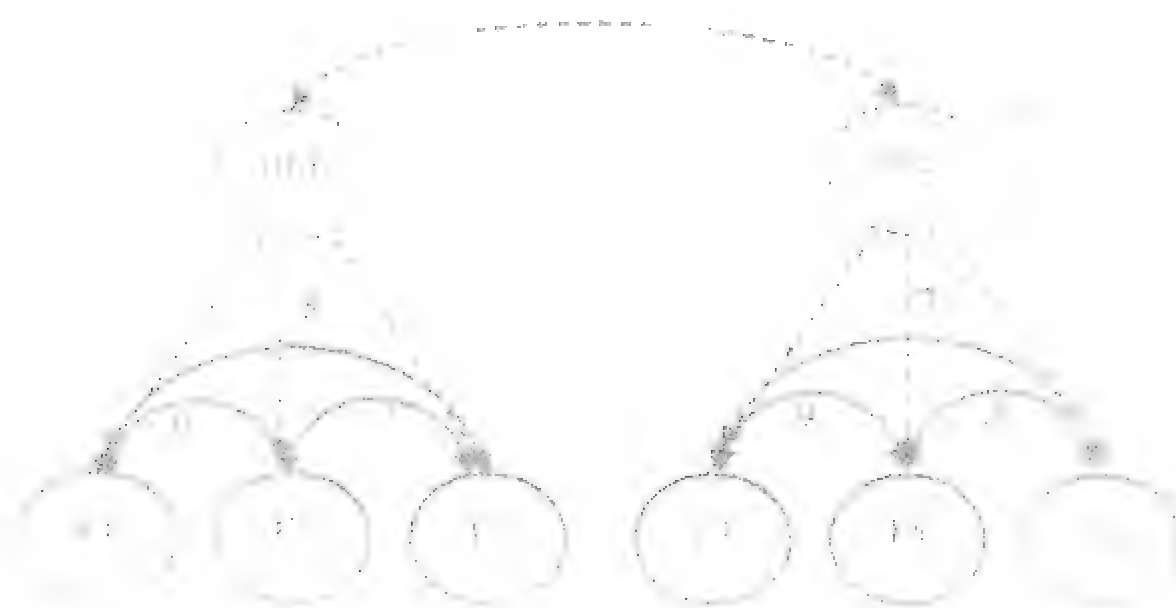
因素分析是当代心理计量领域最重要的统计分析方法之一。其最大的特色是可以处理潜在变量的估计问题，因此深受社会与行为科学家的喜爱。然而，就是因为因素分析法所处理的是社会与行为科学领域最为棘手的抽象构念的问题，所以传统上探索性因素分析法的使用，即存在许多必须小心处理的细节（例如变量相关程度的检测、样本规模大小等）。在操作过程中，牵涉到不少人为主观判断的决策历程（例如因素个数的决定、变量的取舍、因素的命名等），如果研究者没有清楚地交代决策的原则与标准，外人将无法得知因素分析的结果是如何获得，难免给予他人暗箱作业的感觉。

到了 SEM 发展成熟的今天，除了传统的探索性因素分析，研究者多了一项新的因素分析工具，也就是验证性因素分析，得以进行抽象构念的探讨。虽然 CFA 增强了因素分析的理论逻辑的合理性，在应用上比 EFA 增加了许多的弹性（例如允许多维假设的因素分析、提供多重的因素模型检测指数、便捷的模式修饰程序等），但是，因素分析是否能够正确、可靠地被操作，以获得客观、有效的结果，仍是威胁因素分析在学术上的贡献的最关键因素。

Nunnally 与 Bernstein (1994) 提醒研究者，因素分析法不论其功能有多么强大，所得到的数据不论有多么丰富，但是它始终是数学与统计所堆砌的产物，并掺杂着许多缺乏明确客观机制的人为操纵。因此，在检视这些分析成果的同时，研究者仍须不时地回头反省这些统计程序与客观事实之间的关系，并时时检查是否过度解释各项数据或不当操纵分析的过程，藉以满足研究者个人的需求，而非忠诚于科学客观的程序与结论。换句话说，因素分析只是一个手段而非目的。它带领我们深入人类行为现象的抽象世界，最后，还是要倚赖科学的准则以及我们的聪明睿智才能找到出口，勾勒出现实世界的真实面貌。否则，就好比水能载舟亦能覆舟，因素分析能够带领我们找到真理、看到真相，却也可能让我们迷失在人类行为复杂现象与诡谲的统计技术所交织的五里雾中。其中的奥妙与深意，不言而喻。

第六章

高阶验证性因素分析



一般而言，CFA 测量模型所处理的问题是测量变量与潜在变量的关系。利用一组测量变量实际测得的共变结构，抽离出适当的潜在因素，用以检测研究者所提出的假设模型是否合宜、拟合，这些潜在因素直接由测量变量抽离得出，称之为初阶因素（first-order factors）。如果一个测量模型有多个初阶潜在因素，因素之间的共变关系可以加以估计，计算出两两因素相关系数来反映潜在变量之间的关系强度。此时，潜在因素之间存在的是平行的相关关系，各个初阶因素并无特定的隶属结构关系，称之为初阶验证性因素分析（first-order CFA）。如果研究者认为在初阶潜在因素之间，存在有共同的、更高阶的潜在因素，称为阶层验证性因素分析（hierarchical CFA, HCFA），如图 6.1 所示。

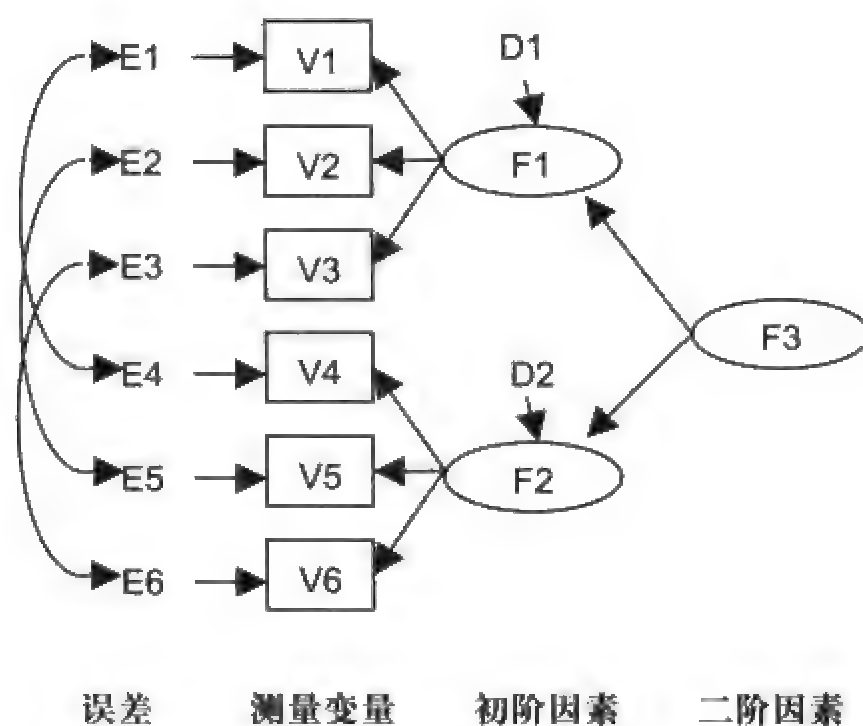


图 6.1 二阶的阶层验证性因素分析图

第一节 高阶验证性因素分析的概念

一、高阶验证性因素分析的模型界定

对于一个测量模式，若研究者认为在初阶潜在因素之间，存在有共同的、更高阶的潜在因素（亦即因素的因素），此种测量模型称为阶层测量模型（hierarchical measurement model），如图 6.1 所示。影响初阶因素的共同影响源称为高阶因素（higher-order factor）。初阶因素之上的一层潜在因素称为二阶因素（second-order factor）（图 6.1 中的 F3），如果还有第三层因素，则称为三阶因素（third-order factor），依此类推。

高阶验证性因素分析的适用时机可以从两方面检视。第一是理论的观点，如果研究者所关心的议题，在理论或概念上可能存在一个高阶因素时，即可利用 HCFA 程序加以检测。例如一项关于创造力的研究中，研究者基于一般学者所采用的多向度理论观点，纳入了图形与语文两种类型的认知性创造测量（包括流畅力、变通力、独创力与精进力四个指标）与情意性创造测量（包括冒险性、好奇心、想象性与挑战性四个子向度），此时共有图形创造力（F1）、语文创造力（F2）、情意创造力（F3）三个初阶因素。研究者还可以假设这三个初阶因素受到一个共同的一般创造力所决定，因而构成了一个高阶验证性因素模型。

再例如自我概念的测量，Marsh 与 Hocevar（1985）发展了一个多向度自我概念工具评估年轻人的自我状态。初阶因素包括了语文、数理、一般性学业、父母、同性、异性、外表、体能等十项自我概念向度。这十个初阶因素背后则受到学业自我概念与非学业自我概念两个高阶因素的影响（详细分析程序可参考邱皓政，2003 的研究实例）。

另一个考虑的层次是实证数据的需要与技术上的考虑。如果研究者从 CFA 模型中观察到初阶因素之间具有一致的高相关时，如图 6.2 的 F1 到 F3 因素的高相关，以及 F4 到 F6 因素的高相关，即可进行高阶因素模型分析，比较 CFA 与 HCFA 模型的拟合度优劣，决定是否初阶因素间具有更高阶的共同影响因素（HF1 与 HF2）。

如果模型比较的结果发现 HCFA 模型更能代表初阶因素的关系，研究者即可利用此一模型在后续的应用上，或是将初阶因素改以组合分数简化成观察变量，将 HCFA 模型降级成 CFA，此一方式的优点是可以减低测量模式的复杂度，提高模式

的简效性，有利于研究者去解释这些潜在因素之间的关系。尤其适合于当研究样本较小的情况下，利用此方式可以提高模型收敛的能力，得到较佳的估计结果。

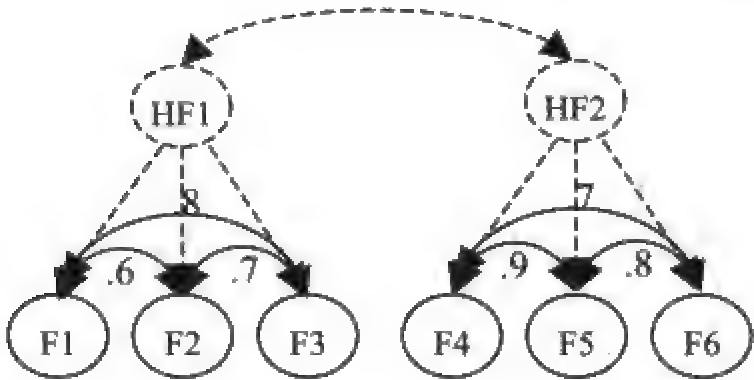


图 6.2 可能带有高阶因素的测量模型

利用阶层化的测量结构，并藉由验证性因素分析进行实证检验，使得研究者可以提出更多的测量模型来相互比较，提高了分析的弹性，尤其在一些行为科学研究，因素之间具有繁复的结构关系，SEM 的分析技术提供了这类弹性检验的可能性。

二、高阶验证性因素分析的模型界定

高阶测量模型又称为全 Y 模型 (all-Y model)，因为模型中的高阶因素被设定为影响初阶因素的外源变量，初阶因素为内生变量，因而所有的观察变量必须被设定为内生潜在变量的观察变量（即 Y 变量），作为外源变量的高阶因素没有任何观察变量（亦即无 X 变量）。一个完整的 SEM 模型由下列三个方程式所组成：

$$y = A_y \eta + \varepsilon \tag{6.1}$$

$$x = A_x \xi + \delta \tag{6.2}$$

$$\eta = B \eta + \Gamma \xi + \zeta \tag{6.3}$$

由于 HCFA 模型中没有 X 变量, 因此公式 (6.2) 即不需要存在, 潜在变量 ξ 由内生潜在变量定义之。换言之, HCFA 模型中仅需估计 Λ_y 即可, 但须增加对于 Γ (Gamma) 矩阵的估计藉以定义潜在变量 ξ 。而且由于初阶因素的相关被高阶因素所取代, 因此 B (Beta) 矩阵为 0 矩阵。

进一步的, 也因为 HCFA 模型中有公式 6.3 的结构参数需要估计, 因此模型中除了测量残差 (ε) 以外, 也会产生新的估计残差 (ς) (即图 6.1 中的 D1 与 D2)。因此一个 HCFA 模型仅需下列两个方程式来决定:

$$y = \Lambda_y \eta + \varepsilon \quad (6.4)$$

$$\eta = \Gamma \xi + \varsigma \quad (6.5)$$

其中公式 (6.5) 中的外源潜在变量的方差协方差矩阵 Φ 中的对角线 (亦即 ξ 变量的方差) 可以设定为 1.0, 共变量可以自由估计。此时 Γ 矩阵即不需要设定任何一个参数为 1.0 作为高阶因素 (ξ 变量) 的量尺化参数。

值得注意的是, 基于模型识别原则, 形成高阶因素的初阶因素数目不能低于三, 否则将造成识别不足问题。因此 HCFA 模型界定的一个简单算法是每一个高阶因素必须被三个或以上的初阶因素所定义。

HCFA 的特色之一是较 CFA 为简效。当形成高阶因素的初阶因素越多时, HCFA 会比原来的 CFA 越简单 (所需估计的参数越少)。诚如先前所言, HCFA 的特色是最高阶的因素作为外源潜在变量, 并没有所属的观察变量, ξ 变量系由低阶的因素估计得出, 因此低阶因素的相关将作为高阶因素的定义的基础。如果有三个带有相关的初阶因素 (有三个协方差参数), 改以一个高阶因素取代这三个初阶因素的相关时, 自由度不变, 亦即是一个等值模型, 模型拟合度会完全相同。但如果是四个带有相关的初阶因素 (有 $C_2^4 = 6$ 个协方差参数), 改以一个高阶因素来取代这四个初阶因素的相关时, 所估计的参数会减少 2, 自由度增加 2, 但卡方值会增加。

HCFA 的简效性优点, 在初阶因素数目为 3 时并不存在, 此时 HCFA 与 CFA 模型的差异主要是理论层次上的考虑。但当初阶因素数目大于 3 时, 两种模型的比较除了理论上的考虑, 还有统计上的意义: 当初阶因素数目大于 3 或更多时, HCFA 模型的简效性优点则趋凸显, 但所付出的代价是模型拟合变差。只要 HCFA 模型的

拟合度没有显著地比 CFA 模型差时,我们即可依据简效法则,接受 HCFA 模型为较佳模型。

第二节 高阶验证性因素分析的实例

一、创造力理论的因素效度检证

为了说明 HCFA 在理论模型验证上的应用,我们在此以一个典型的心理学理论研究——创造力研究作为范例。

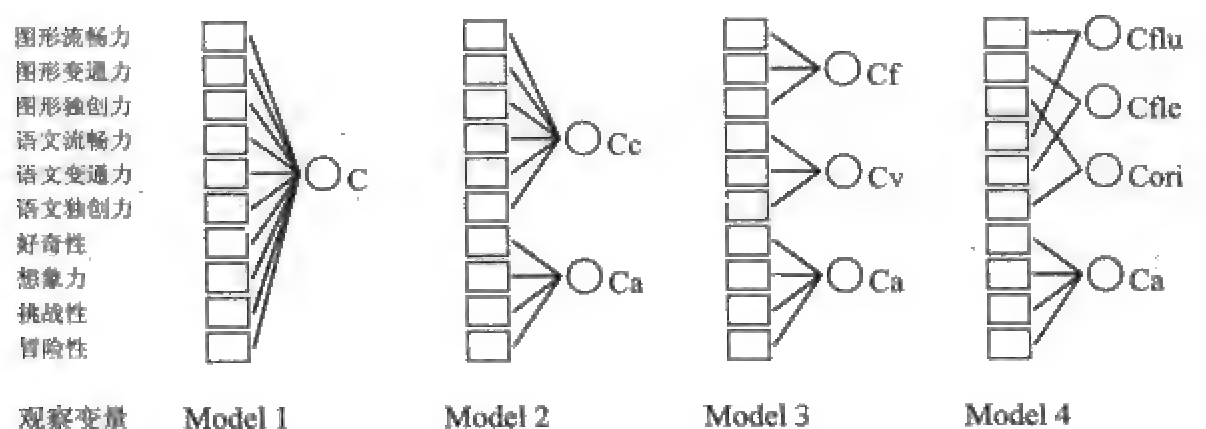
一般而言,研究者普遍认为创造力不是用单一指标可以测量得到的概念,而是呈现多向度、多层面的概念。从创造力的本质来看,人类的创造力可以区分为认知与非认知层面,Guilford (1962) 认为认知层面的创造力可以依据认知历程所处理的材料的不同区分为图形与语文创造力等不同形式。认知创造能力由四种不同历程组成:流畅力、变通力、独创力与精进力。在非认知层面方面,Williams 认为高创造力者应具有好奇性、想象性、挑战性及冒险性等四种情意成分,因此提出了四成分的创造性情意的理论概念。综合这两位学者的观点,可以得到一个创造力的认知与情意二成分理论模型,将创造力区分为认知层面与情意层面,而认知层面又区分为图形与语文两类型,这些创造力概念可以分别由不同的测量方式得到观察分数,再以 CFA 及 HCFA 模型检验该理论模型的适合度。

在测量上,Guilford 的认知创造力可以由 Torrance 创造思考测验测得(吴静吉,高泉丰,王敬仁以及丁兴祥,1981),针对语文与图形两种形式,测得受试者的流畅力、变通力、独创力与精进力分数(由于精进力计分不易,本范例不予纳入)。情意创造力的概念,则可以由 Williams 于 1980 所编制的创造力评量组测验中的扩散性情意测验(Test of Divergent Feeling),测得个体在好奇性、想象性、挑战性及冒险性等四项创造性情意强弱程度(王木荣,林幸台,1987)。

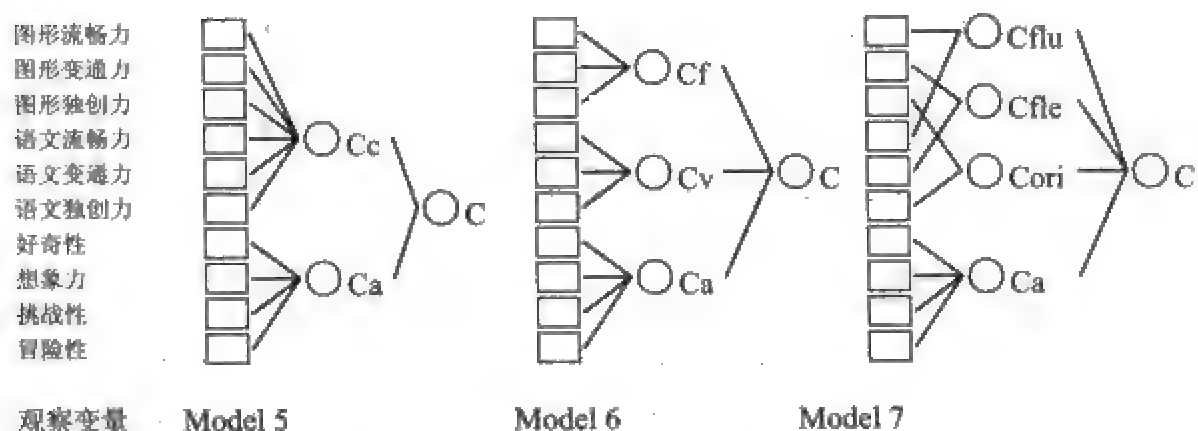
二、CFA 与 HCFA 的操作

(一) 模型界定

基于前述的创造力理论，可以设定一个创造力测量的 CFA 模型。图形流畅力、图形变通力、图形独创力、语文流畅力、语文变通力、语文独创力、好奇心、想象力、挑战性及冒险性等十个创造力分数为十个观察（测量）变量，根据不同的概念基础，我们可以利用不同的假设，提出不同的组合形式。分述如下：



(a) 初阶假设模型



(b) 高阶假设模型（二阶）

图 6.3 各种创造力测量分数的假设模型

模型一：十个创造力分数受到同一个潜在变量的影响，称为一般性创造力因素(C)，这是CFA模型中，具有最少的潜在变量(因素)的一个最简单初阶模型(Model 1)。

模型二：基于认知与非认知的区分，六个认知分数受到同一个潜在变量的影响(一般性创造力认知因素，Cc)，而Affect力受到另一个潜在变量的影响(一般性创造力情意因素，Ca)(Model 2)。

模型三，认知成分的六个得分受到两种认知形式的影响(图形 Ccf 与语文 Ccv)(Model 3)。

模型四：认知成分的六个得分由三个成分组成(流畅力 Cflu、变通力 Cfle、独创力 Cori)(Model 4)。

上述四个模型如图 6.3 中的 Model 1 至 Model 4。这四个模型均为初阶模型，也就是十个观察变量的背后受到一群潜在因素的影响，但是潜在因素的背后，并没有更高阶的因素。如果考虑这些初阶因素背后还有更高阶的因素，就必须以高阶验证性因素分析检验下列几项模型。

模型五：模型二的两个初阶因素 Cc 与 Ca 背后受到一个二阶因素(second-order factor)一般性创造力因素(C)的影响(Model 5)。

模型六：模型三的两个认知初阶因素(图形 Ccf 与语文 Ccv)与情意因素(Ca)背后受到一个二阶的(一般创造力因素，C)的影响(Model 6)。

模型七：模型四的三个认知初阶因素(流畅力 Cflu、变通力 Cfle、独创力 Cori)与情意因素(Ca)背后受到一个二阶的(一般性创造力因素，C)的影响(Model 7)。

后面三个模型均为高阶模型，因为初阶因素的背后均存在更高阶的因素，如图 6.3 的 Model 5 至 Model 7 所示，这几个 CFA 即为 HCFA。从图 6.4 的 CFA 概念路径图，可以算出测量数据数为 $(10 \times 11) / 2 = 55$ (DP=55)。Model 3 为 CFA 模型(图 6.4)，Model 6 为 HCFA 模型(图 6.5)，两者的模型界定条件差异并不大。

从图 6.4 的路径图可以看出，初阶模型的因素之间仅具有相关而没有高阶潜在变量的设定。而一旦在初阶因素之上再增加一个高阶因素之时，图 6.5 则同时增加了三个干扰项(ζ_1 、 ζ_2 、 ζ_3)，也就是以外源潜在变量(ζ_1 二阶因素)解释三个内生潜在变量(η_1 、 η_2 、 η_3)的解释残差。其他的各参数均维持一致。值得注意的是，为了使高阶因素得以

估计, 必须将高阶因素的方差设为 1, 或是把定义高阶因素的Gamma系数之中的其中一个设定为 1 (在 LISREL 的默认是外源潜在变量的方差为 1)。

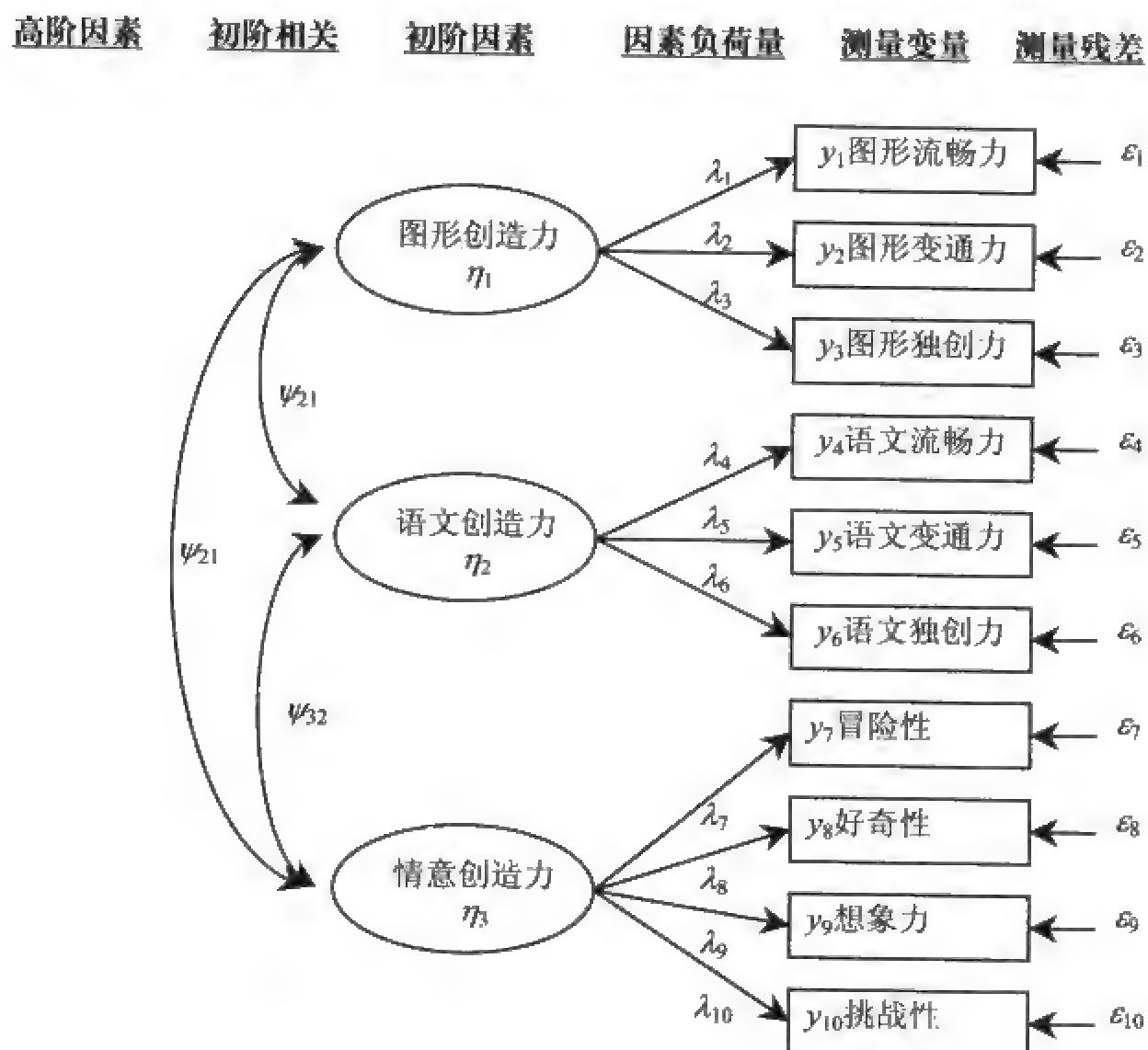


图 6.4 CFA 模型 (Model 3) 的假设模型路径图

HCFA 的执行方法与步骤, 除了语法指令的撰写与 CFA 稍有不同之外, 其余皆无异。因此, 以下仅针对初阶模型与高阶模型各举一个假设模型进行分析示范, 然后以摘要表的形式进行各模型的比较。

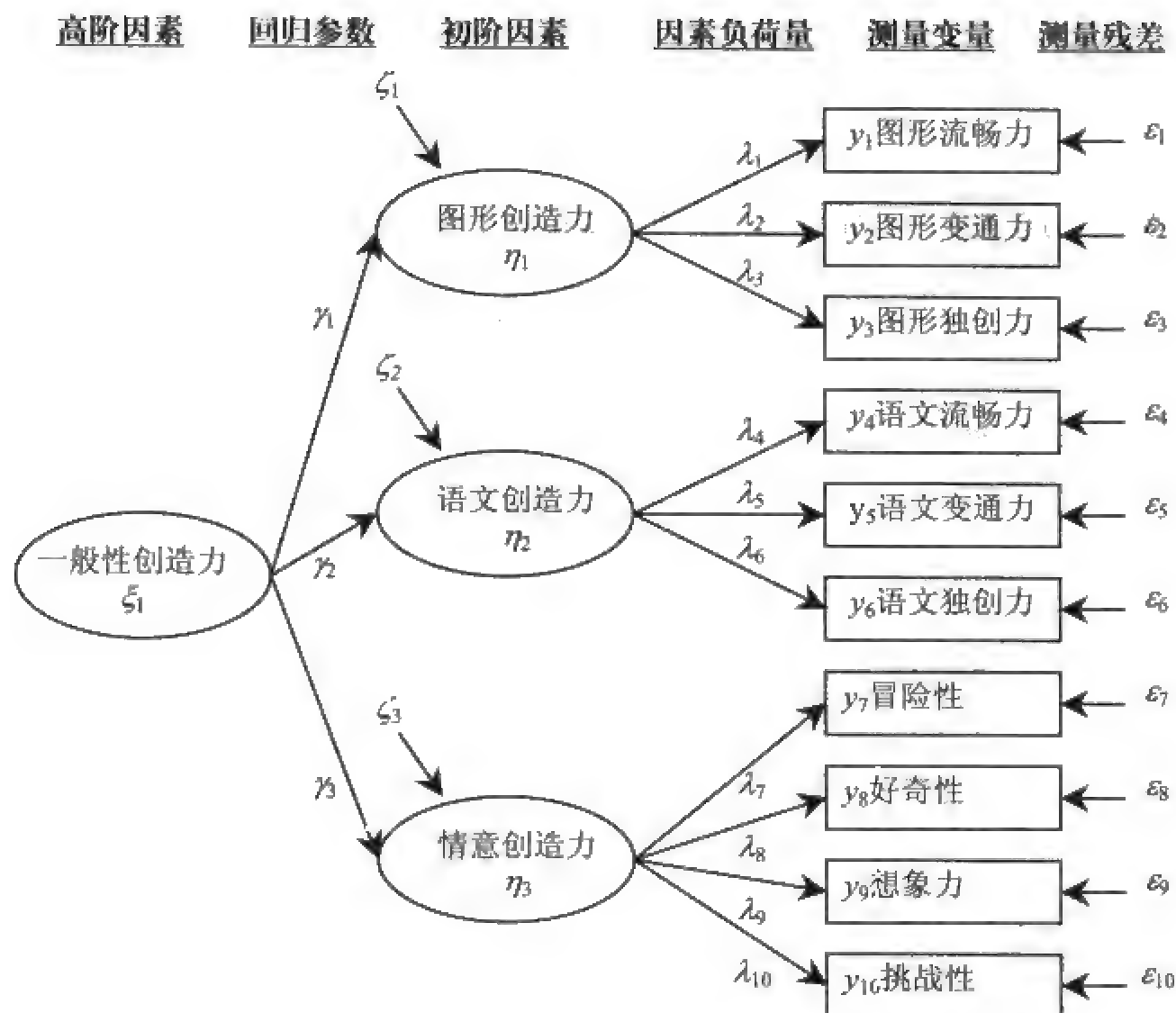


图 6.5 HCFA 模型 (Model 6) 的假设模型路径图

(二) LISREL 语法

本范例以原始数据库作为输入数据,原始数据库储存于 ch6.dat, Model 3 (CFA) 与 Model 6 (HCFA) 的 LISREL 语法列举表 6.1。

表 6.1 的 LISREL 语法,前面的数据读入指令完全相同。变量数目为 18,样本数目 804,数据以原始数据形式读入。从 MO 指令开始则有所变化。Model 3 与 Model 6 的不同在于 Model 指令。Model 3 为了估计因素间的相关,必须增加 PS=DI,FR 指令,或在语法中直接释放因素间的共变使其得以估计。PS=DI,FR 则说明初阶因素(各

内生潜在变量)被高阶因素解释后所剩余的独特变异 PSI 矩阵中, 对角线与下三角部分(初阶因素的残差方差与协方差)被自由估计。

表 6.1 Model 3 与 Model 6 的 LISREL 语法(ch6m3.ls8 与 ch6m6.ls8)

```

Title ch6 Higher-Oder CFA using LISREL syntax Model 3
DATA  NI=10 NO=804
RA FI='ch6.dat' FO
(10F3.0)
LA;  FF1 FF2 FF3 VF1 VF2 VF3 PF1 PF2 PF3 PF4
MO  NY=10 NE=3 PS=SY,FR
LE;  Figual Verbal Affect
FR  LY 2 1 LY 3 1 LY 5 2 LY 6 2 LY 8 3 LY 9 3 LY 10 3
VA 1 LY 1 1 LY 4 2 LY 7 3
PD
OUTPUT SE TV SS MI

```

```

Title ch6 Higher-Oder CFA using LISREL syntax Model 6
DATA  NI=10 NO=804
RA FI='ch6.dat' FO
(10F3.0)
LA;  FF1 FF2 FF3 VF1 VF2 VF3 PF1 PF2 PF3 PF4
MO  NY=10 NE=3 NK=1 LY=FU,FI GA=FU,FR
LE;  Figual Verbal Affect
LK;  Crea
FR  LY 2 1 LY 3 1 LY 4 2 LY 6 2 LY 8 3 LY 9 3 LY 10 3
VA 1 LY 1 1 LY 5 2 LY 7 3
PD
OUTPUT SS SC MI

```

至于高阶模型则增加了一个外源潜在变量(NK=1), 此一新增的高阶潜在变量(二阶因素)由三个初阶因素所定义, 因此必须增加 GA=FU,FR 指令。GA=FU,FR 说明了高阶因素(外源潜在变量)与初阶因素(各内生潜在变量)的 GAMMA 矩阵所有的结构参数被自由估计。

(三) SIMPLIS 语法

以 SIMPLIS 语法进行前述模型的估计，语法列于表 6.2。CFA 与 HCFA 的语法差异，在于 HCFA 模型中增加了初阶与高阶因素的关系设定：Figural Verbal Affect=Crea，并将高阶因素的方差设为 1.0：Set the Variance of crea to 1。除此之外则无不同（但是要记得在潜在变量名称中多增列高阶因素的名字：Latent Variables: Figural Verbal Affect Crea）。

另外，由于本范例是读取原始数据，如果是以固定格式来读取原始数据，在数据库（Ch6.dat）档案中的第一行必须标明数据格式：10F3.0，否则 SIMPLIS 无法识别。

表 6.2 Model 3 与 Model 6 的 SIMPLIS 语法(ch6m3.spl 与 ch6m6.spl)

Ch6 Higher-Order CFA using SIMPLIS syntax model 3
Observed Variables: FF1 FF2 FF3 VF1 VF2 VF3 PF1 PF2 PF3 PF4
Raw Data from File ch6.dat
Sample Size = 804
Latent Variables: Figural Verbal Affect
Relationships:
FF1-FF3=Figural
VF1-VF3=Verbal
PF1-PF4=Affect
Path Diagram
LISREL Output
End of Program
Ch6 Higher-Order CFA using SIMPLIS syntax model 6
Observed Variables: FF1 FF2 FF3 VF1 VF2 VF3 PF1 PF2 PF3 PF4
Raw Data from File ch6.dat
Sample Size = 804
Latent Variables: Figural Verbal Affect Crea
Relationships:
FF1 FF2 FF3=Figural
VF2=1*Verbal
VF1 VF3=Verbal
PF1-PF4=Affect
Figural Verbal Affect=Crea
Set the Variance of crea to 1
Path Diagram
LISREL Output
End of Program

(四) 结果报告与说明 (以 HCFA 为例)

1. 模型界定

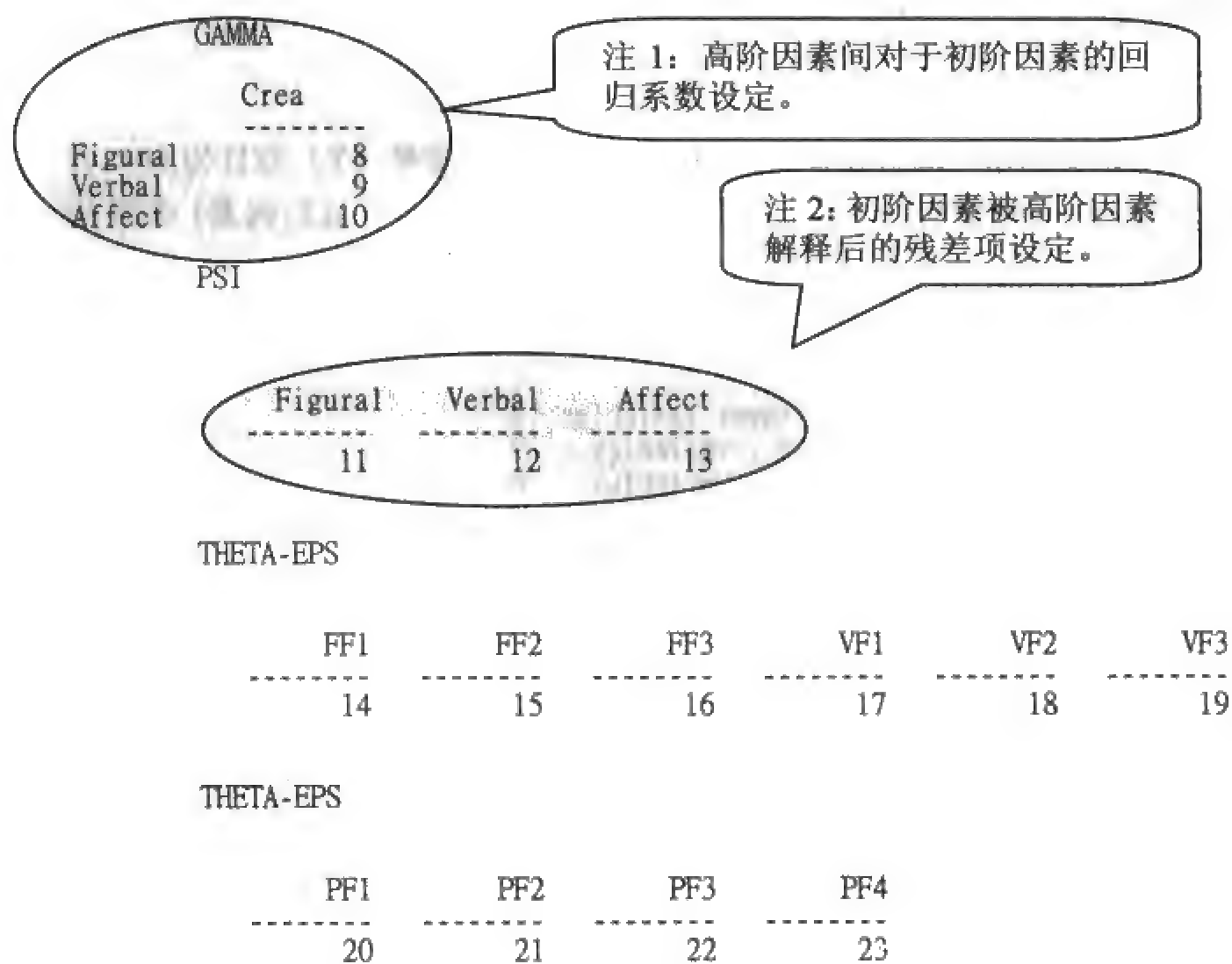
Model 6 的报表指出, 测量变量共计 10 个, 内生测量变量 (Y) 数目为 10, 内生潜在变量 (ETA; 初阶因素) 数目为 3, 外源潜在变量 (KSI; 高阶因素) 数目为 1, 样本数为 804。

```
Number of Input Variables 10
Number of Y - Variables 10
Number of X - Variables 0
Number of ETA - Variables 3
Number of KSI - Variables 1
Number of Observations 804
```

接着列出了 Model 6 所进行估计的 23 个参数, 分别是因素载荷 7 个 (1 至 7)、结构参数 3 个 (8 至 10) (注 1)、初阶因素被高阶因素解释剩余的残差变异量 3 个 (11 至 13) (注 2)、以及观察变量测量误差 10 个 (14 至 23)。注 1 与注 2 所列为高阶因素所影响的参数。

Parameter Specifications

	LAMBDA-Y		
	Figural	Verbal	Affect
FF1	0	0	0
FF2	1	0	0
FF3	2	0	0
VF1	0	3	0
VF2	0	0	0
VF3	0	4	0
PF1	0	0	0
PF2	0	0	5
PF3	0	0	6
PF4	0	0	7



2. 参数估计结果

Model 6 所得到的结果指出, LISREL 总计进行了 73 次的迭代才完成了所有的参数估计。但是在范例二中, 没有高阶设定的 Model 3 仅需要 4 次迭代即完成了参数估计, 然而, Model 6 的各初阶参数的因素载荷估计数, 则与 Model 3 完全相同 (注 3), 显见高阶因素的设定提高了整体模型估计上的难度, 而非个别参数估计的本身。结果输出如下:

higher-order Confirmatory Factor Analysis #3 Model 6

Number of Iterations = 73

LISREL Estimates (Maximum Likelihood)

LAMBDA-Y

	Figural	Verbal	Affect
FF1	1.00	---	---
FF2	0.63 (0.01) 50.00	---	---
FF3	1.61 (0.03) 49.15	---	---

... (略)

注 3: 参数估计结果与显著性检验与 Model 3 完全相同, 表示高阶因素的设定并不影响低阶因素的参数估计。

GAMMA

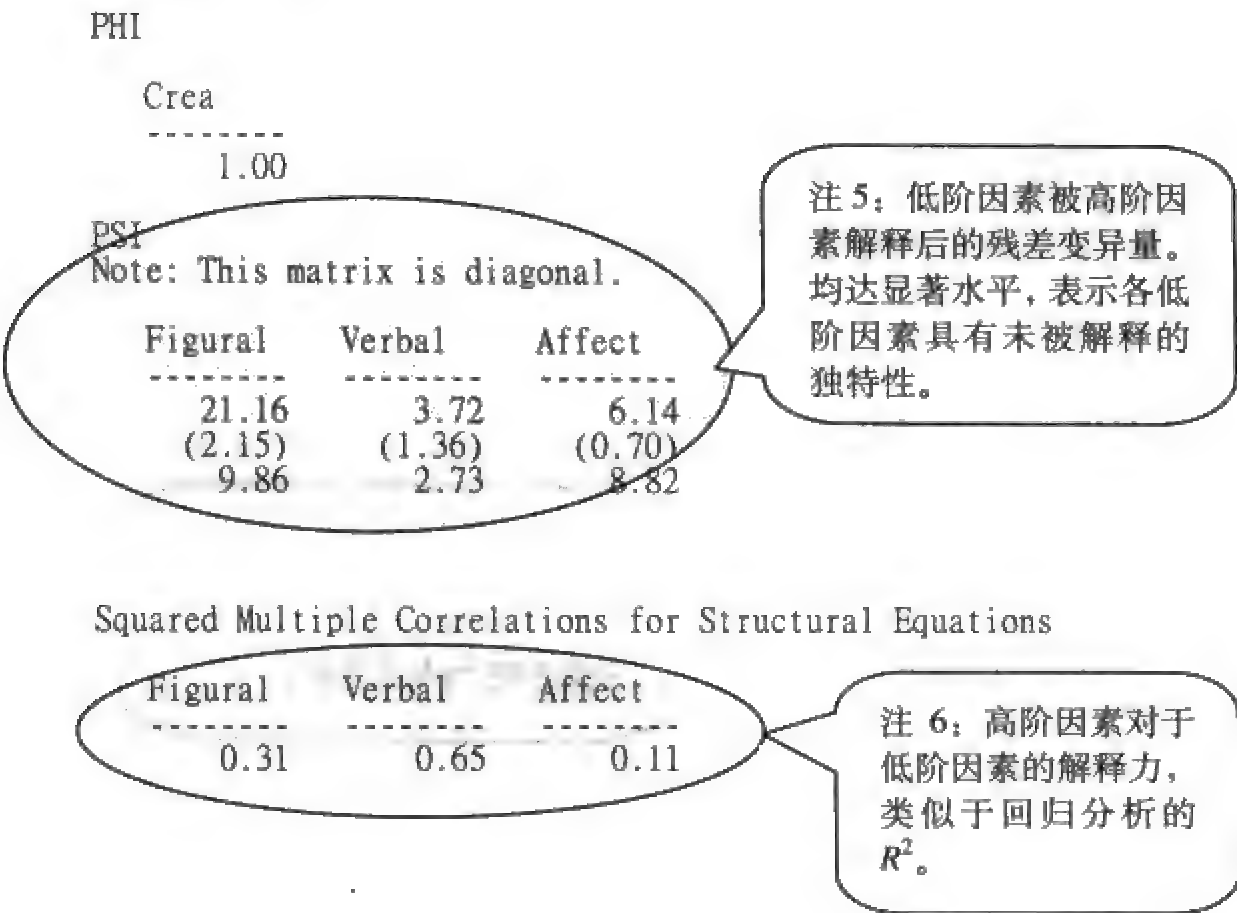
Crea

Figural	3.06 (0.35) 8.75
Verbal	2.62 (0.28) 9.39
Affect	0.88 (0.14) 6.36

Covariance Matrix of ETA and KSI

	Figural	Verbal	Affect	Crea
Figural	30.52			
Verbal	8.02	10.59		
Affect	2.70	2.31	6.92	
Crea	3.06	2.62	0.88	1.00

注 4: 潜在变量的方差与协方差参数估计结果。包括了高阶因素与初阶因素。



上述参数估计结果显示，所有的参数估计数均具有显著的意义。潜在变量的方差与协方差的估计数（ETA 与 KSI 矩阵）虽然被列出（注 4），但是并未提供显著性检验，因为这些参数仅是作为 GAMMA 矩阵估计的基础。

高阶因素对初阶因素的结构参数取代了 ETA 与 KSI 矩阵，造成 ETA 与 KSI 矩阵的估计数不被继续处理，但是却造成了 PSI 需被检验，也就是各低阶因素被解释后的残差矩阵（PSI）。数据显示这些残差项均达显著水平（注 5），表示低阶因素具有高阶因素之外的独特变异（uniqueness）。而报表亦指出，三个初阶因素可以被解释的变异量分别为 31%、65%、11%，显示高阶因素（一般性创造力）对于语文认知创造力的解释较强，对于情意创造力的解释力最弱（注 6）。

紧接着是测量变量的测量残差估计数（THETA-EPS）与显著性检验。所有的数据与范例二的 Model 3 的估计值与显著性完全相同，证明了高阶因素的设定并不会影响初阶因素与测量变量之间的关系。

3. 模型拟合度分析

Model 6 的模型拟合指数与其他模型的模型拟合数据列于表 6.3。从模型拟合指

数来看, Model 3 与 6 是相当理想的模型, 但两者模型拟合度数据完全相同, 显示两者去反映真实观测资料的优劣程度一样, 无法比较。

表 6.3 各模型拟合度比较表

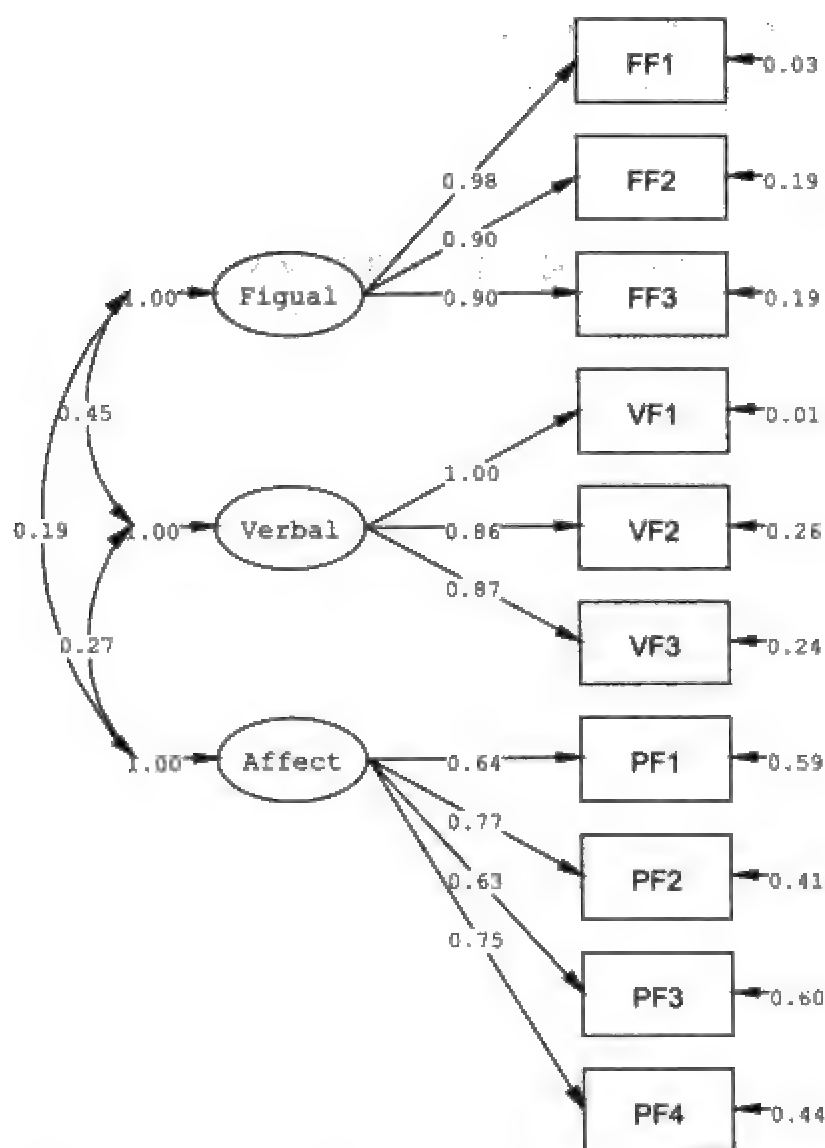
Model	Chi-square	df	RMSEA	NNFI	CFI	SRMR
Model 1	2914.87	35	.32	.39	.52	.22
Model 2	1699.68	34	.25	.56	.67	.16
Model 3	86.75	32	.046	.99	.99	.02
Model 4	1533.16	29	.25	.52	.69	.12
Model 5	1700.50	33	.25	.54	.67	.67
Model 6	86.75	32	.046	.99	.99	.99
Model 7	2513.22	31	.32	.17	.43	.43

注: Model 2 为 Model 5 的初阶模型, Model 3 为 Model 6 的初阶模型, Model 4 为 Model 7 的初阶模型。

(五) 分析结果与参数报告

先前的数据提供了模型估计与适合度的数据, 但是在最后阶段, 我们必须决定何者是我们所欲取用的数据。一般而言, 应选择模型拟合度较佳的模型, 作为我们最终模型。但是, 如果两个模型拟合度相当, 卡方值的差异未达显著水平, 即无法识别何者的拟合度较佳。

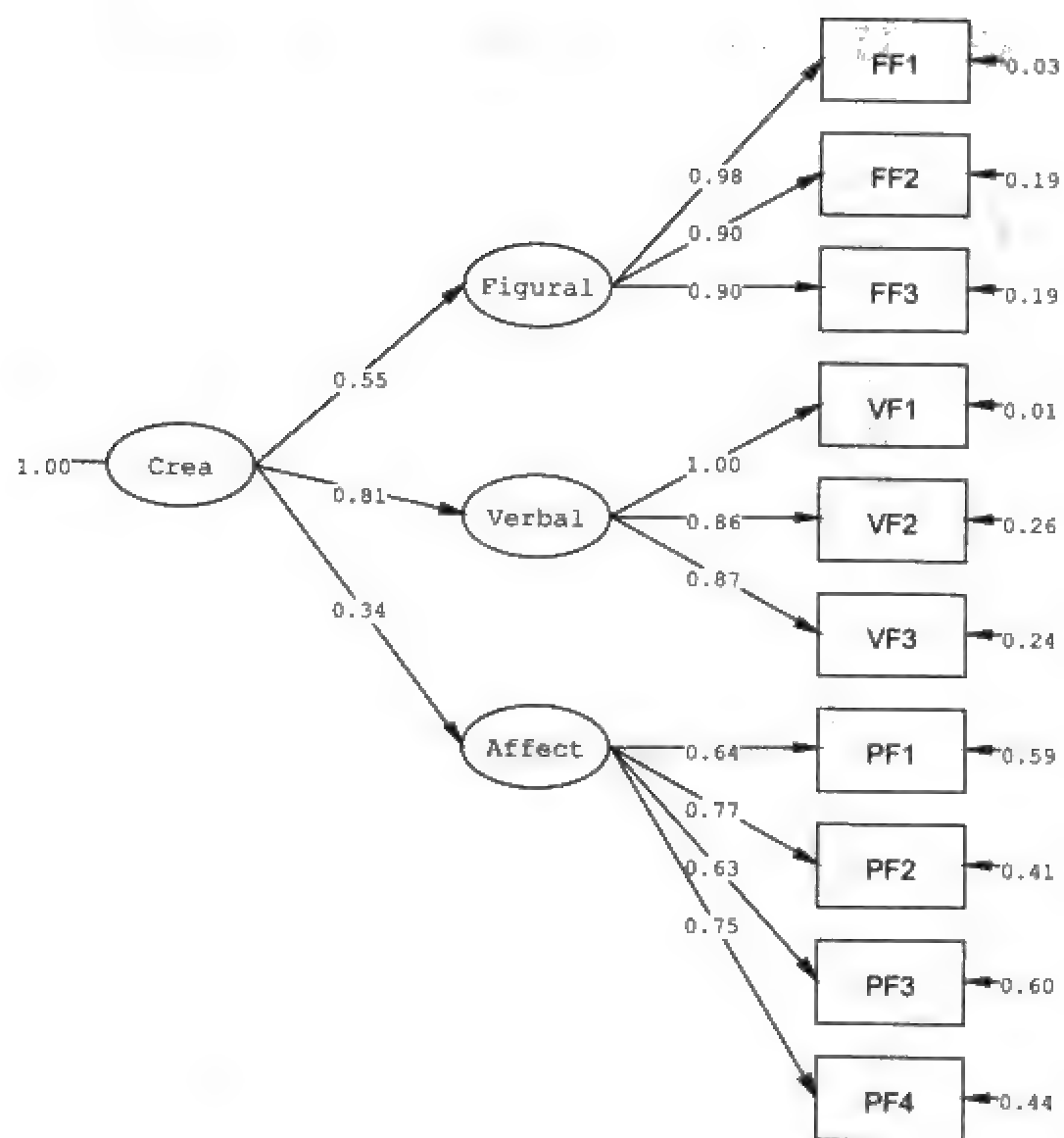
Model 6 与 Model 3 自由度完全相同, 拟合度数据完全相同, 证实了先前所讨论的两者为等值模型 (equivalent model)。两个模型的拟合优劣无法比较。Model 6 仅在高阶因素设定上有所变化, 并未影响到测量变量的设定, 因此模型整体拟合情形没有变化。如此一来, 我们即无法利用模型拟合度进行模型比较, 无法看出 Model 3 与 Model 6 的优劣性。最后则将 Model 3 与 Model 6 的终解的所有参数估计详列于图 6.6 与图 6.7, 读者可以自行比较其间的异同。



Chi-Square=86.75, df=32, P-value=0.00000, RMSEA=0.046

图 6.6 Model 3 标准化终解路径图

就 HCFA 而言，最重要的系数莫过于高阶因素的因素载荷，该参数反映了高阶因素对于初阶因素的解释力，列于 Gamma 矩阵中。报表数值指出，高阶因素（一般创造力）对于语文认知性创造力的影响最大（ $r=.81$ ），其次为图形认知性创造力（ $r=.55$ ），最后是情意创造力（ $r=.34$ ）。在 Model 3 当中，这些高阶因素的因素载荷，是由初阶因素之间的相关来表示，分别为图形认知与语文认知的.45、图形认知与情意认知的.19 以及语文与情意创造力的.27。 r 系数与相关系数的性质相差甚大，同时数据的意义与解读的方式也大不相同，因此究竟 Model 3 与 Model 6 何者可以用来反映创造力的理论因素效度，除了可以依循简约原则判断，也应该审视数据所提供的信息，配合文献与学理的讨论，得到最合理的论证。



Chi-Square=86.75, df=32, P-value=0.00000, RMSEA=0.046

图 6.7 Model 6 标准化终解路径图

第三节 AMOS 的高阶验证性因素分析

一、模型界定

依照创造力理论的假设图，绘制图形，图 6.8 为执行完毕之后的标准化最终解路径图（以 Model6 为例）。

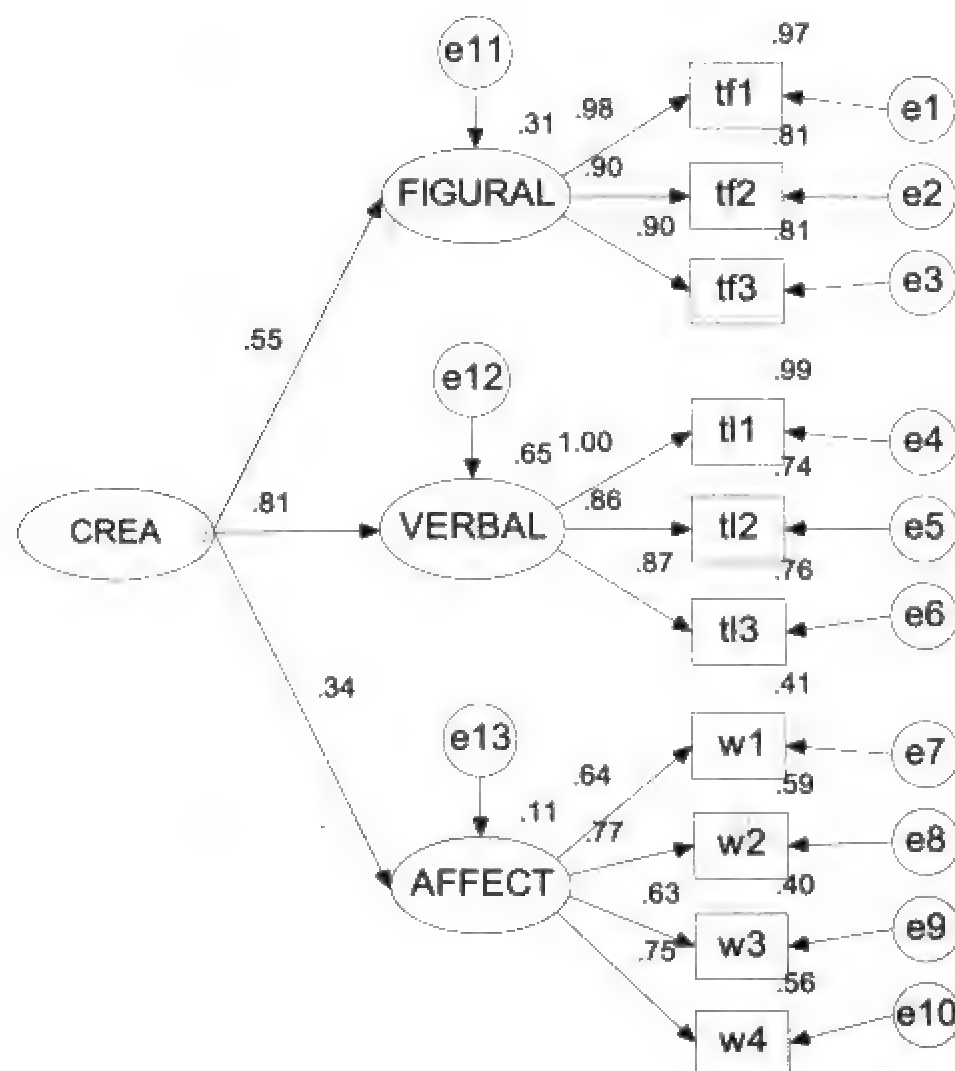
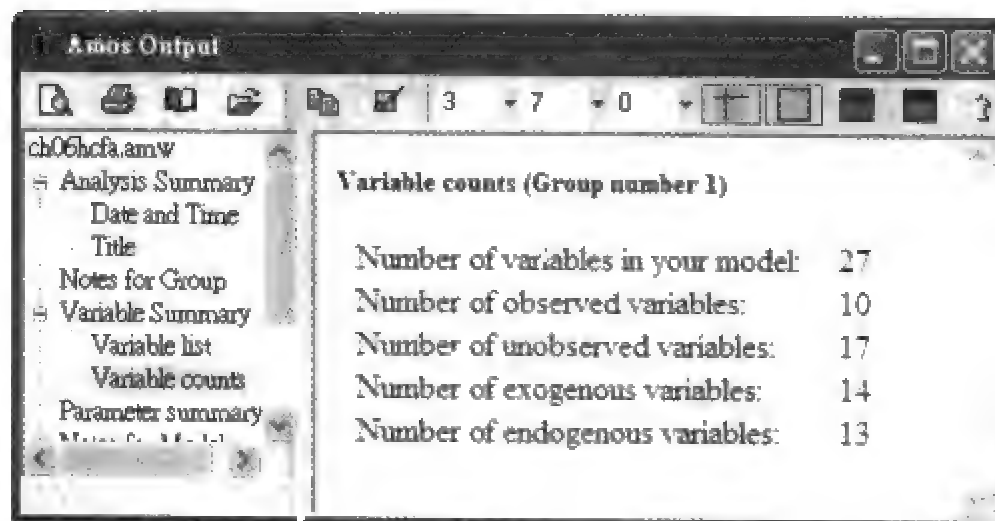


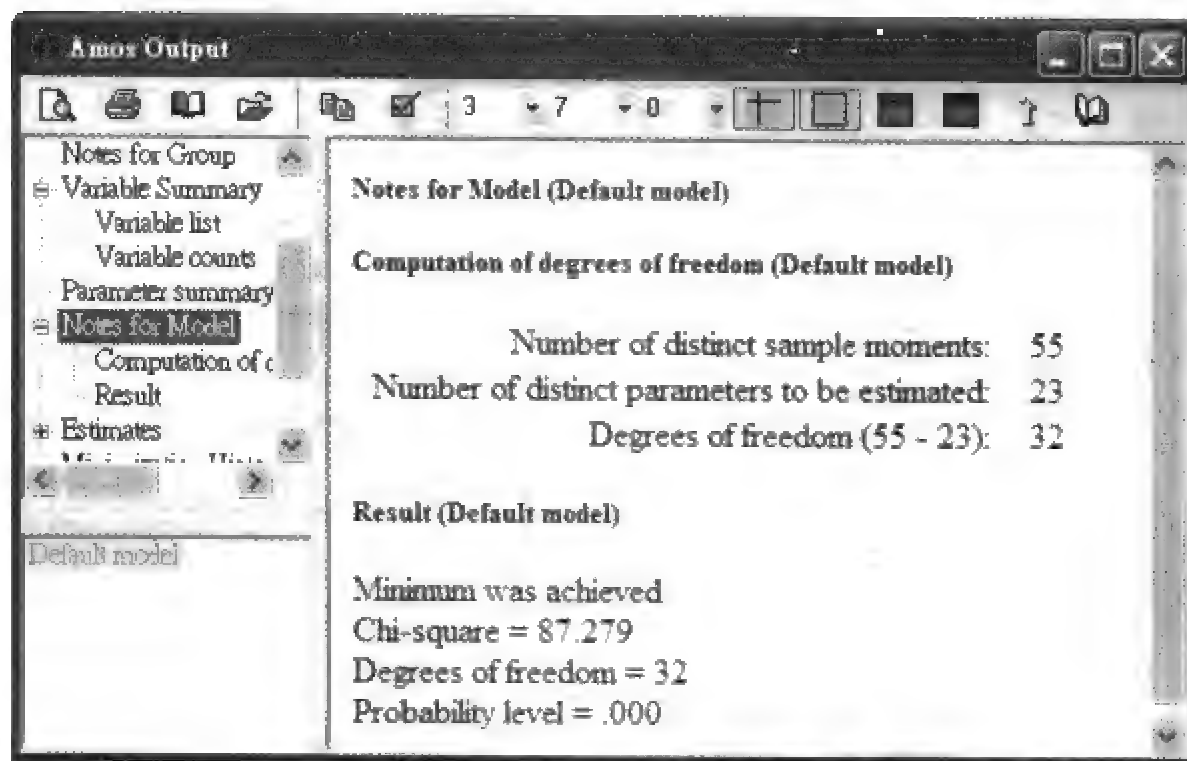
图 6.8 以 AMOS 绘制的 HCFA 概念图与参数估计结果

二、AMOS 的报表解读

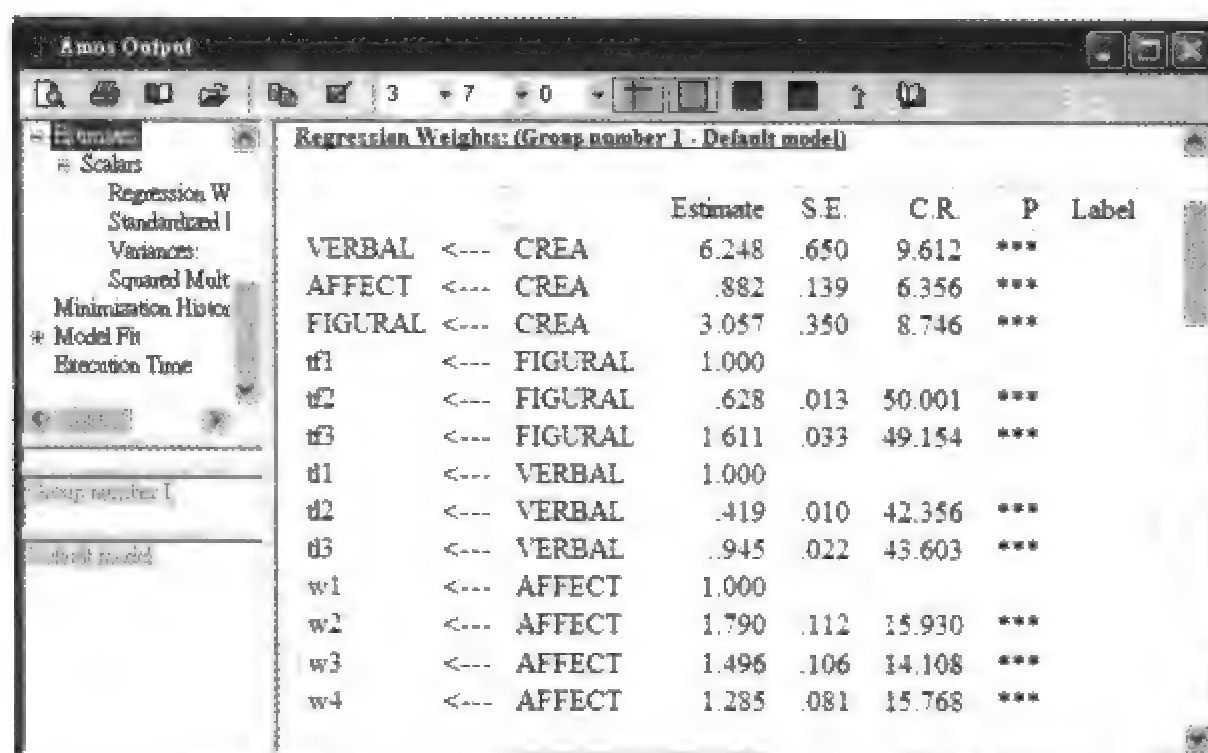
变量的摘要显示模型中的所有变量，包含观察变量（内生变量）与非观察变量（外源变量），观察变量有 10 个，非观察变量有 17 个，共计 27 个变量。



模式的纪录显示模型的测量数据数为 55 个，有 23 个参数被估计，因此自由度为 32。在模式的纪录也能看到卡方值(87.28)与显著性(.000)。

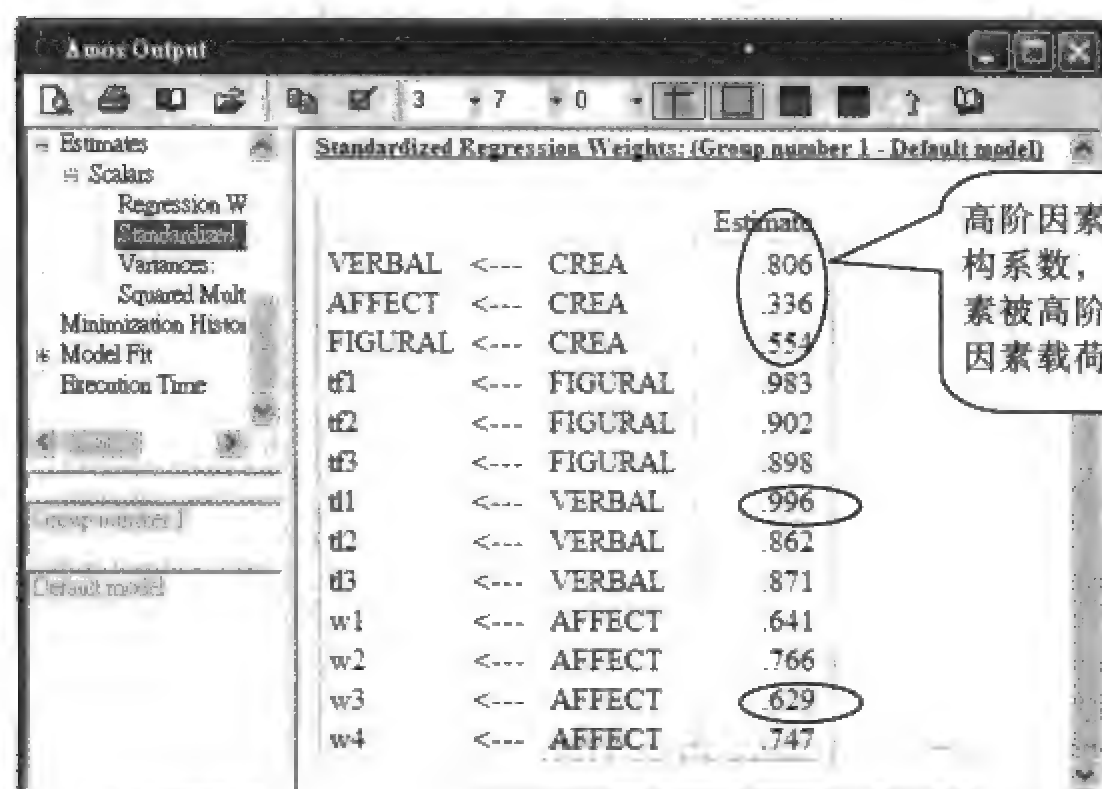


从参数估计结果的报表数据可以得知,所有的参数值均达显著,因素载荷以 t11 的.996 最高, w3 的.629 最低。高阶因素的标准化结构系数,亦即初阶因素被高阶因素解释的因素载荷,一般创造力 (CREA) 对图形创造力 (FIGURAL) 解释的因素载荷为.554,一般创造力对语文创造力 (VERBAL) 为.806,一般创造力对情意创造力 (AFFECT) 为.336。



Regression Weights: (Group number 1 - Default model)

		Estimate	S.E.	C.R.	P	Label
VERBAL	<--- CREA	6.248	.650	9.612	***	
AFFECT	<--- CREA	.882	.139	6.356	***	
FIGURAL	<--- CREA	3.057	.350	8.746	***	
t1	<--- FIGURAL	1.000				
t2	<--- FIGURAL	.628	.013	50.001	***	
t3	<--- FIGURAL	1.611	.033	49.154	***	
d1	<--- VERBAL	1.000				
d2	<--- VERBAL	.419	.010	42.356	***	
d3	<--- VERBAL	.945	.022	43.603	***	
w1	<--- AFFECT	1.000				
w2	<--- AFFECT	1.790	.112	15.930	***	
w3	<--- AFFECT	1.496	.106	14.108	***	
w4	<--- AFFECT	1.285	.081	15.768	***	



Standardized Regression Weights: (Group number 1 - Default model)

		Estimate
VERBAL	<--- CREA	.806
AFFECT	<--- CREA	.336
FIGURAL	<--- CREA	.554
t1	<--- FIGURAL	.983
t2	<--- FIGURAL	.902
t3	<--- FIGURAL	.898
d1	<--- VERBAL	.996
d2	<--- VERBAL	.862
d3	<--- VERBAL	.871
w1	<--- AFFECT	.641
w2	<--- AFFECT	.766
w3	<--- AFFECT	.629
w4	<--- AFFECT	.747

高阶因素的标准化结构系数,亦即初阶因素被高阶因素解释的因素载荷量。

三、模型拟合度分析

从前面的模型的纪录中显示拟合卡方值 (Chi-square) 为 87.279 ($p=.000$)，从模型拟合度摘要表中也显示 $CMIN/DF$ 为 2.727。平均概似平方误根系数 $RMSEA=.046$ ； $SRMR=.020$ ， TLI ($NNFI$)、 NFI 、 CFI 、 GFI 皆大于 .90，显示 Model6 具有理想的拟合度。

Model Fit Summary	
CMIN	
Model	NPAR CMIN DF P CMIN/DF
Default model	23 87.279 32 .000 2.727
Saturated model	55 .000 0
Independence model	10 5868.627 45 .000 130.414
RMR, GFI	
Model	RMR GFI AGFI PGFI
Default model	.802 .979 .964 .570
Saturated model	.000 1.000
Independence model	15.510 .398 .264 .325
Baseline Comparisons	
Model	NFI Delta1 RFI rho1 IFI Delta2 TLI rho2 CFI
Default model	.985 .979 .991 .987 .991
Saturated model	1.000 1.000 1.000
Independence model	.000 .000 .000 .000 .000

Standardized RMR	
Default model	Standardized RMR = .0204

第四节 Mplus 的高阶验证性因素分析

一、Mplus语法

```

TITLE:   ch6 Higher-Order CFA
DATA:   FILE IS ch6.dat;
VARIABLE: NAMES ARE FF1 FF2 FF3 VF1 VF2 VF3 PF1 PF2 PF3 PF4;
MODEL:
  Figural by FF1-FF3;
  Verbal  by VF1-VF3;
  Affect  by PF1-PF4;
  Crea    by Figural* Verbal Affect;
  Crea@1;
OUTPUT:
STANDARDIZED;

```

设定模型

各初阶因素的第一条参数的 λ 自动
预设为 1.0。

设定高阶因素

高阶因素 Crea 由三个初阶因素所
定义。三个参数自由估计。

高阶因素的
方差设为
1.0。

二、Mplus估计结果

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value	87.388
Degrees of Freedom	32
P-Value	0.0000

Chi-Square Test of Model Fit for the Baseline Model

Value	5875.935
Degrees of Freedom	45
P-Value	0.0000

CFI/TLI

CFI	0.991
TLI	0.987

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.046	
90 Percent C.I.	0.035	0.058
Probability RMSEA \leq .05	0.677	

SRMR (Standardized Root Mean Square Residual)

Value	0.019
-------	-------

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
FIGURAL BY				
FF1	0.983	0.005	214.120	0.000
FF2	0.902	0.008	117.103	0.000
FF3	0.898	0.008	112.618	0.000
VERBAL BY				
VF2	0.862	0.010	83.738	0.000
VF1	0.996	0.006	179.773	0.000
VF3	0.871	0.010	89.037	0.000
AFFECT BY				
PF1	0.641	0.026	24.574	0.000
PF2	0.766	0.022	34.895	0.000
PF3	0.629	0.027	23.668	0.000
PF4	0.747	0.023	33.152	0.000
CREA BY				
FIGURAL	0.554	0.060	9.283	0.000
VERBAL	0.806	0.080	10.090	0.000
AFFECT	0.336	0.048	6.938	0.000
Intercepts				
FF1	2.094	0.063	33.231	0.000
..... (略)				
Variances				
CREA	1.000	0.000	999.000	999.000
Residual Variances				
FF1	0.034	0.009	3.771	0.000
FF2	0.186	0.014	13.359	0.000
FF3	0.193	0.014	13.494	0.000
VF1	0.008	0.011	0.721	0.471
VF2	0.256	0.018	14.419	0.000
VF3	0.242	0.017	14.202	0.000
PF1	0.589	0.033	17.614	0.000
PF2	0.413	0.034	12.281	0.000
PF3	0.605	0.033	18.108	0.000
PF4	0.442	0.034	13.136	0.000
FIGURAL	0.693	0.066	10.495	0.000
VERBAL	0.351	0.129	2.729	0.006
AFFECT	0.887	0.032	27.314	0.000

第七章

路径分析

路径分析 (path analysis) 是一种将观察变量间的关系以模型化的方式进行分析的一种统计技术。在实际分析操作上, 可以利用统计软件进行多次的多元回归分析, 得到所有的路径系数并加以组合。然而, 自从 SEM 发展以来, 路径分析可以轻易地在 SEM 加以模组与检验, 因此路径分析逐渐改以 SEM 的方式来处理, 称为结构方程模型取向 (SEM approach) 的路径分析。其主要特色是可以利用测量变量间的共变情形, 同时 (simultaneously) 估计模型当中所有的参数, 并配合研究者所提出的特定假设模型或竞争模型, 检验理论模型与观察数据的適切性, 找出最佳的模型。

第一节 路径分析的基本概念

一、相关与因果

2008 年 5 月 12 日下午 2 时 28 分左右, 中国四川省发生 8.0 级的大地震。震中汶川首当其冲, 死伤无数。如果人类能够预知地震, 那么数万人的生命即得以挽救。

然而, 根据相关报道, 四川绵竹市西南镇檀木村在地震前两日曾出现大规模蟾蜍迁徙, 数十万蟾蜍走上马路。另据报道, 4 月 26 日早上 7 时, 湖北下村坝中一个直径约百米水深数十米的水塘约 8 万立方米的蓄水全部消失, 并伴有轰鸣声。犹有甚者, 地震发生前半小时, 邻近的甘肃有民众拍到天空出现五颜六色的美丽彩云霞光, 认为这是地震要发生之前的“地震云”。一时之间, 似乎地震是可以预知的, 而是我们疏于关心探究。但是, 地震真的可以预知吗?

确实曾有科学家指出, 地震要发生之前, 天空中会出现一种白色的带状白云, 称之为地震云。其原理是因为地球本身是一个大磁场, 当地震来临前, “地壳变动”影响了“地底磁场”, 进而影响“云层中电子排列”而形成地震云。这套说法如果合并其他异象, 可以绘制成图 7.1 的关系。

图 7.1 的这个因果关系图如果获得证实, 那么, 或许人们无法掌握因为地壳变动 (X) 所产生的地震 (Y) ($X \rightarrow Y$), 但可以透过天象的观测 (地震云 $M1$ 、动物迁徙 $M2$) 的中介变量 (mediation) 预知地震 ($M1 \rightarrow Y$ 、 $M2 \rightarrow Y$) 而逃过一劫。虽然科学界一直持续相关研究, 但却始终无法证明天空中云彩的变化以及动物的迁徙和地

震的关联机制。演变成芸芸百姓不断汇报一些与地震“相关”的讯息，但科学家始终不愿做出“因果”的结论。为什么呢？难道真的因为科学家缺乏大智慧，又或者是研究者并没能善用路径分析去掌握天象与地震之间的因果关系。

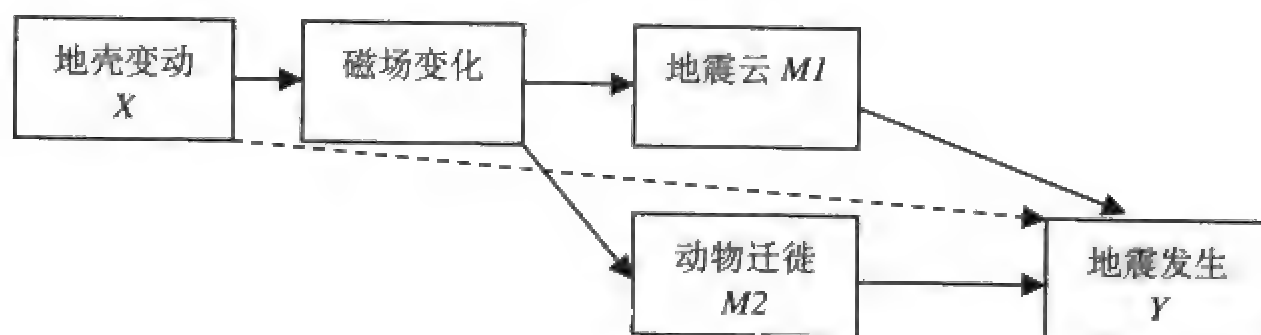


图 7.1 大地震前的预兆

除了自然科学家关心因果，心理学家、社会学家、经济学家、父母老师甚至于政治人物，无人不关心。观察相关、掌握因果，是学术工作者最大的梦想，但是却也最遥不可及。

1921 年，学经济的 Wright 首度提出了“路径分析”的概念，利用回归方程式联结图 7.1 当中的 X （自变量，independent variable; IV）、 M （中介变量，mediator; Me）、 Y （因变量，dependent variable; DV），建立一个统计模型解释一组有关联的变量之间背后的因果关系，以推得经济上的因果结论（causality）。

简单来说，路径模型的目的是在解释为何一组变量会有关系？它们是如何彼此相互影响？因此背后的数学逻辑以及分析的基本素材是观察变量之间的共变结构，这也是路径分析与其他所有的 SEM 技术相通之处。共变结构反映的是一组变量之间的多元化共变性，而个别两两变量的共变强度可以利用标准化共变（也就是相关系数）描述两者线性关系的强度。

几乎所有的基本统计教科书上都会提醒大家，相关不等于因果（correlation does not prove causality），但在路径分析中，这条警语被修改为相关蕴涵了因果关系；变量之间可以存在因果关系；透过适当的程序与分析，配合假设检验，相关可以推导出因果结论（causal inference can be strongly corroborated by correlation）。

Meehl 与 Waller (2002) 指出, 路径分析虽然可以检验因果论证的存在, 但是对于谁为因谁为果, 却需要更为严谨的检验 (correlation evidences causation but does not immediately reveal what causes what, and how much, p. 284)。虽然 Wright 发明了一套用以检测因果关系的有力工具, 但如果把某一个研究的分析结果视为特定变量因果关系的唯一证据, 而没有多方检验其真实性, 将可能造成偏差的结论¹。循此, 也正说明了正确学习与运用路径分析的重要性。

二、第三变量与中介效应

虽然因果关系是建立在 IV (X) 对 DV (Y) 的预测关系上 (以 $X \rightarrow Y$ 表示)。但是路径分析不同于传统的回归分析, 是因为在 $X \rightarrow Y$ 的关系中存在着至少一个中介变量, 形成一个 $X \rightarrow Me \rightarrow Y$ 的中介效应 (mediation effect)。Mackinnon (2008) 指出一般所谓的路径分析, 即是一连串中介效应的组合所形成的复杂模型, 路径分析的关键就是中介变量以及中介变量之间的复杂关系。换言之, 路径分析是整体模式的总称, 中介关系则是路径分析的核心, 串连中介关系的元素就是变量的共变结构。因此, 当代盛行的结构方程模型恰能提供路径分析一个最佳的分析网络 (详细内容请参考 Mackinnon 的近作)。

根据韦伯字典的定义, 中介 (mediate) 一词是指位居中间的位置 (to be in an intermediate position or sides), 或是联系两个人或物的中间传递者 (to be an intermediary or conciliator between persons or sides)。因此一个中介关系的存在, 起始于一组不对称的 IV 与 DV 的双极关系, 亦即存在着一个有意义的 $X \rightarrow Y$ 双变量效应 (two-variable effects)。进一步的, 由于第三个变量 (Z) 的引入, 形成一组待检验的三变量效应 (three-variable effects)。此一 Z 未必就是中介变量, 而可能以其他的形式存在并发生不同类型的影响 (Mackinnon, 2008)。Mackinnon (2008) 整理了第三变量的各种可能形态, 例如干扰变量 (confounder), 意指因为它同时与 IV 与 DV 有关, 而可能影响 IV 与 DV 之间关系的混淆 (obscures) 与突出 (accentuates) (Meinert, 1986)。IV 与 DV 之间的关系若是受到干扰变量的影响呈现减弱、弱化时, 称为虚假关系 (spurious relationship)。相反的, IV 与 DV 之间的关系可能因为

¹ 有关因果关系的论证, 一直以来即是一个相当有争议的议题, 读者可以参考 Cliff (1983)、Freedman (1987, 1997)、Clogg & Haritou (1997) 的论文。

干扰变量的存在而加强时, 干扰变量称为压抑变量 (suppressor variable) (Conger, 1974; Mackinnon, Krull, & Lockwood, 2000)。不论是弱化或强化, 干扰变量对于原来的 IV 与 DV 的关系会产生一种扭曲与混淆的效果, 甚至于把 IV 与 DV 的关系的正负方向完全逆转, 此时该干扰变量称之为逆转变量 (distorter variable) (Rosenberg, 1968), 一般在学术界多以统计控制策略来排除干扰变量的影响。在实际研究中, 对于模型中某个变量, 排除了该变量的干扰效应时, 普遍被称之为控制变量 (control variable)。

另一种第三变量称之为共变量 (covariate), 是指当 Z 与 X 都能解释 Y, 但是 X 与 Z 之间关系不明显时。一般来说, 干扰变量与共变量的区分不甚明显, 但干扰变量被视为不利于解释 $X \rightarrow Y$ 的效应而欲除之的骚扰变量 (nuisance variables), 而共变量则是有助于理解 $X \rightarrow Y$ 效应的解释变量, 因此又被称为共存变量 (concomitant variable), 因为它对于模型解释力的影响明显, 不能轻易地从模型中移出 (Cochran, 1957; Smith, 1957)。

另外两种第三变量的存在形式可以说是社会科学领域重要的议题, 亦即中介者与调节者。在一个统计模型中, 中介者 (mediator, Me) 扮演 IV 与 DV 中继的角色, 调节者 (moderator, Mo) 则让 $X \rightarrow Y$ 的效应有条件地产生变化。换言之, IV 与 Mo 会对 DV 产生交互作用 (interaction effect), 使得在调节变量的不同水平下, $X \rightarrow Y$ 的效应产生系统性变化 (Sharma, Durand, & Gur-Arie, 1981)。

诚如先前所述, 路径分析是强调变量间影响关系的解释, 因此第三变量的角色主要是以中介变量的形式被检验。当然一个中介模型的第三变量是干扰变量、共变量或调节变量。在本章的讨论中, 则仅以中介变量的角度检验路径模型的分析策略。

三、中介效应的检验

Baron 与 Kenny (1986) 从统计方法学的角度, 提出了中介效应的完整概念体系与检验程序。其中中介效应系利用三个回归方程式检测四个条件是否成立, 如图 7.2 所示。

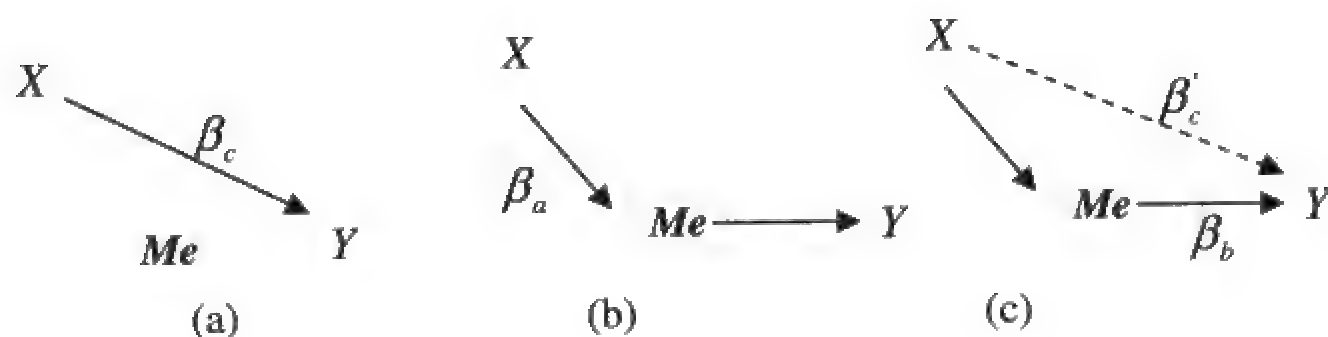


图 7.2 中介效应图示

首先, 中介效应的第一个条件为 β_c 的估计值必须具有统计显著性。当 β_c 的估计值显著, 代表 X 对 Y 有影响, 亦即 Y 变量的变异可以被 X 所解释, 如图 7.2(a) 所示。图 7.2(a) 显示了自变量 X 对因变量 Y 的总效应估计, 此时中介变量 Me 并未进行任何的处理, 方程式如下:

$$Y_i = \alpha_c + \beta_c X_i + \varepsilon_i^c \quad (7.1)$$

中介效应的第二个条件为 β_a 的估计值必须具有统计显著性。当 β_a 的估计值显著, 代表自变量 X 对中介变量 Me 有影响, 亦即 Me 变量的变异可以被自变量 X 所解释, 如图 7.2(b) 所示。图 7.2(b) 显示了中介变量对自变量 X 的回归, 回归方程式如下:

$$Me_i = \alpha_a + \beta_a X_i + \varepsilon_i^a \quad (7.2)$$

第三个条件是在同时考虑自变量 X 与中介变量 M 对于因变量的影响时, 中介变量 Me 必须具有统计显著性。亦即在包含 X 与 Me 变量的多元回归方程式中, β_b 系数必须具有统计显著性, 方可证明中介变量 Me 在排除自变量 X 后仍对因变量 Y 有净影响:

$$Y_i = \alpha_b + \beta_c' X_i + \beta_b' Me_i + \varepsilon_i^b \quad (7.3)$$

上述方程式即为图 7.2 (c) 的表述, 在同时考虑中介变量 Me 与自变量 X 进入回归方程式后, 其中介变量 Me 与自变量 X 独特对 Y 的影响。第四个条件为当控制中介变量 Me 后, 原先的自变量 X 其净效应消失, 亦即 β_c' 的估计值没有到达统计显著水平。

如果上述四项条件完全符合, 亦即 Me 完全中介了 X 对 Y 的效应, 亦即一种完全中介效应 (completed mediation effects); 如果 β_c' 的估计值虽有变化, 但仍具有统计显著性, 若其绝对值小于 β_c' 的估计值, 则称为 Me 部分中介了 X 对 Y 的效应, 亦即部分中介效应 (partial mediation effects)。

在方程式(7.1)与(7.3)中, X 对 Y 的直接效应为 β_c 与 β_c' 的差异, 代表 X 经过 Me 对 Y 的间接效应, 称为中介效应, 可以 $\beta_c - \beta_c'$ 表示。在单一层次的中介效应路径图中, 上述三个方程式的回归系数具有下列关系 (MacKinnon, Warsi, & Dwyer, 1995):

$$\beta_c - \beta_c' = \beta_a \beta_b \quad (7.4)$$

Sobel(1982)也曾推导出样本估计量 $\hat{\beta}_a \hat{\beta}_b$ 标准误的一阶与二阶泰勒展开式估计值:

$$S_{\hat{\beta}_a \hat{\beta}_b}^2 = S_{\hat{\beta}_b}^2 \hat{\beta}_a^2 + S_{\hat{\beta}_a}^2 \hat{\beta}_b^2 \quad (7.5)$$

$$S_{\hat{\beta}_a \hat{\beta}_b}^2 = S_{\hat{\beta}_b}^2 \hat{\beta}_a^2 + S_{\hat{\beta}_a}^2 \hat{\beta}_b^2 + S_{\hat{\beta}_b}^2 S_{\hat{\beta}_a}^2 \quad (7.6)$$

因此, 除了考虑 Baron 与 Kenny(1986)提出的四条件外, 亦可以利用检定 $\hat{\beta}_a \hat{\beta}_b$ 是否显著来判断存在中介效应。Baron 与 Kenny(1986)的四个条件是利用统计技术检测“统计的”中介效应, 然而从方法学的观点来看, 中介效应隐含着因果关系的推论, 必须在 X 对 Me 有因果关系、 Me 对 Y 有因果关系以及 X 对 Y 有因果关系的条件下, 方能做出存在“中介效应”的结论。

第二节 路径分析的模型界定与识别

一、理论先行

路径分析最初由遗传学家 Sewall Wright 于 1921 年所提出,至 20 世纪 60 年代才广泛受到重视的一种用以探讨多重变量之间因果结构模式的统计技术。不仅在心理学、教育或社会学等传统社会科学领域有重要的地位,在商学研究或经济计量学上,路径分析都受到相当的重视,例如在经济学领域,路径分析被称之为联立方程模型(simultaneous equation modeling),很早就被提出讨论。

路径分析的企图在建立因果解释模型,也是社会科学领域得以用来检测因果模式的重要策略。但是正因为因果关系所倚赖的基本前提多半难以确立,使得路径分析的结论也经常遭到质疑。例如因果关系的两个变量应具有明确的前后时间关系,即 X 发生在前, Y 发生在后, X 与 Y 的共变方可以被推导为因果关系。理论上,这种 X 先于 Y 发生的资料仅存在于追踪(longitudinal)研究,但多数的社会与行为科学研究却是以同时性(concurrent)测量来收集各观察变量的数据,也就是变量的数据的发生时间相同。因此,因果关系的假设,有赖研究者提出清楚、合理、明确的逻辑与推理程序说明假设存在的基础,例如提出“性别影响自我概念,但是自我概念不会改变性别”之类的论证,或征引理论与文献之支持确立假设的合宜性与合理性。

另一个重要的问题是如何排除与控制其他变量的影响,使得因果关系的存在具有相当的稳定性与内部关系的有效性。传统上,实验研究的优点即是可以精确控制其他变量,藉以观察 X 变量对于 Y 变量的影响。但是多数的社会与行为科学研究所探讨的变量数目多、关系复杂,许多变量因为基本性质或伦理考虑无法在实验室中操弄,使得实验研究无法执行,大量使用统计控制程序的回归分析、路径分析或 SEM 即成为复杂的共变关系分析的重要替代方案,但同时也是这类技术共同问题的根源。统计控制的程序往往因为不同的处理方式造成数据的变异,在分析过程当中常常因为控制变量的调整造成估计结果的改变,甚至扭转研究的结论。因此,从技术层面追求一个稳定的、具有统计检验力(power)的参数估计程序的路径分析达成真相的发掘,其效果实属有限。根本的解决之道仍是建立在适切的理论基础与严谨的假设建构过程,并时时注意统计技术本身的限制与问题(Cliff, 1983)。

从上述的分析中可以看出,路径分析与其他的 SEM 技术需要强而有力的理论基

础作为统计检验过程的支撑。虽然在技术上,路径分析与 SEM 技术提供了各种应对方案处理各种问题,配合计算机优越的数据处理能力可以协助处理繁复的运算,但是根本的关键取决于知识的基础。深度的文献检阅、扎实的理论内涵与缜密的逻辑推论是提出假设模型必要的前提条件,这些研究执行的重要工作落实是研究成败的关键。

二、模型的建立

路径分析的首要步骤是建立一个有待检验的路径模式 (path model)。模式的建立,除了基于研究者所关心的变量与关系,也必须斟酌相关研究与过去文献的观点,提出一个适当的模型以待验证。但是要如何选择适当的变量与安排适当的假设关系,是路径模型建立的两大挑战。以变量的选择来说,对于某一个因变量 Y (如学业表现) 的影响因素除了研究者所关心的变量 (如动机因素) 之外,可能存在其他重要的变量 (例如智力),在模型中是否纳入重要的其他变量,考验着研究者的智慧。

从技术的观点来看,纳入越多对于因变量具有解释力的自变量,可以有效提升模型的拟合度,但是从研究的观点,纳入过多的变量对于现象的解释不但没有帮助,反而容易造成概念上的混淆。因此,研究者无不希望提出一个精简的模型而能解释最大的变异,太多的讯息反而造成判断的困扰。相反的,有时候研究者对于哪些变量对于因变量具有影响力缺乏足够的讯息,尤其对于一些较为冷门或新兴的议题,过去文献与既有理论可能相当缺乏,造成研究者不知如何选择变量,或是无法提出有力的论证以支持模式的意义。过多或过少的信息,对于模型的建立都带来困扰,最后的解决方案还是有赖研究者自行透过归纳、推理与主观的分析决定。

其次是变量关系的决定,表 7.1 列举了路径模型中可能存在的变量关系,除了 \leftrightarrow 代表没有方向性的相关之外,其他的关系类型都与因果关系有关。相关代表着变量间的共变关系,当研究者获得观察数据之后,任何两个变量的共变可以直接计算出来,所以相关可以说是具体存在而可以估算的变量关系,是整个路径分析的基础。其他所有关系的检验,则建立在研究者的假设之上,需要适当的理论依据作为基础。

综上所述,建构一个路径模型时,对于变量的选择与关系的拟定具有高度艺术性。除了藉由理论文献的检视与个人主观的判断推导之外,尚必须考虑模型可识别性等技术层次的问题,避免造成参数无法估计的窘境。

表 7.1 路径分析的各种变量关系

变量与符号	代表意义	关系类型
$X \leftrightarrow Y$	相关 Correlation	X 与 Y 为共变关系
$X \rightarrow Y$	单向因果关系 direct causal effect	X 对 Y 为直接效应
$X \rightarrow Y_1 \rightarrow Y_2$	单向因果关系 direct causal effect	X 对 Y_1 为直接效应, X 对 Y_2 为间接效应, Y_1 为中介变量
$X \rightleftarrows Y$	回溯因果关系 reciprocal causal effect	X 与 Y 互为直接效应, X 与 Y 具有回馈循环效果
$Y_1 \rightarrow Y_2 \rightarrow Y_3 \rightarrow Y_1$	循环因果关系 indirect loop effect	Y_1 对 Y_2 、 Y_2 对 Y_3 、 Y_3 对 Y_1 均为直接效应, Y_1 、 Y_2 与 Y_3 为间接回馈循环效果

三、递归模型与非递归模型

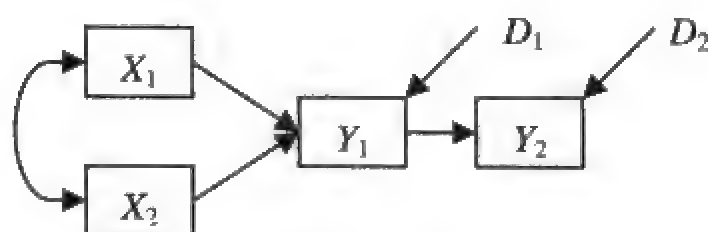
(一) 模型的定义

藉由变量关系的安排, 路径分析有两种不同的基本类型, 递归 (或不可逆) 模型与非递归 (可逆) 模型, 这两种模型的区别主要在于是否具有回溯性或循环因果关系 (见表 7.1)。

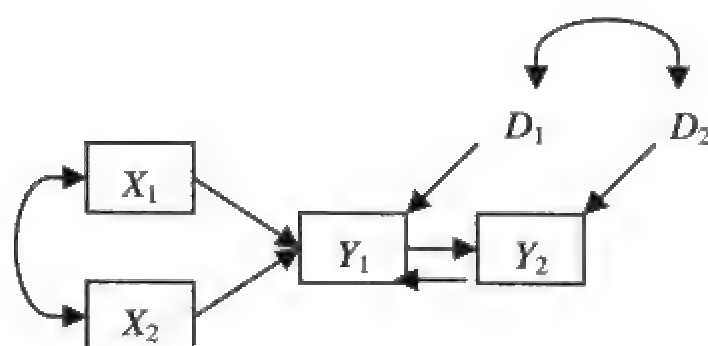
一般回归模型由于有明确的 IV 与 DV 两种变量, IV 对 DV 的影响有一定的次序位阶, 所以不会有可逆或不可逆问题。但是在路径分析中, 变量数目至少会超过三个以上, 因此“因”变量与“果”变量之间关系就相对复杂 (这也就是为什么路径分析不再使用 IV 与 DV 来描述预测变量与被预测变量, 而改以外源或内生变量来描述变量间关系, 因为有时被预测变量却会变成另一变量的预测变量)。

假如现在有四个变量, 两个是作为解释他人的自变量 (以 X_1 与 X_2 表示, 也即是

外源变量), 另外两个是被 X_1 与 X_2 解释的因变量 (以 Y_1 与 Y_2 表示, 也即是内生变量)。 Y_1 与 Y_2 之所以被称为内生变量, 是因为这两者一定会被其他变量解释, 亦即 Y_1 与 Y_2 的变异量将被任何其他变量解释, 解释不足之处称为干扰 (disturbance), 以回归分析的术语来说, 就是指估计后的残差方差 (error variance of the estimation)。相对的, X_1 与 X_2 之所以称之为外源变量, 是因为这两者只专门去解释他人, 他们自己的变异量不被任何变量所解释, 因此 X_1 与 X_2 不会有干扰项。



(a) 递归模型



(b) 非递归模型

图 7.3 路径分析的递归模型与非递归模型

在回归分析中, 残差方差是因变量不能被自变量解释的部分, 计算方法是以 1 减去步骤二所计算出来的 R^2 再乘以因变量的方差, 得到非标准化残差方差 (若不乘以内生变量方差则是标准化的方差):

$$Var_D = (1 - R^2) \times S_y^2 \quad (7.7)$$

值得注意的是，虽然 Y_1 与 Y_2 一定会被其他变量所解释，但这里所谓的“其他变量”除了指外源变量（ X_1 与 X_2 ）之外，也有可能是内生变量。换言之， Y_1 除了被 X_1 与 X_2 解释之外，也可以去解释 Y_2 ；同样的， Y_2 除了被 X_1 与 X_2 解释之外，也可以去解释 Y_1 。然而，为了维系模型的稳定性与可解释性， Y_1 若作为 Y_2 的解释变量时，不宜“同时”作为被 Y_2 解释的内生变量，此时，即是路径分析所谓的递归模型（如图 7.3a 所示）。在递归模型中，任何 IV 与 DV 的配对都是不可逆的。

有时，有些研究无可避免会碰触到 Y_1 作为 Y_2 解释变量， Y_2 又返回作为 Y_1 的解释变量，例如，经济学中的供给与需求之间的关系，或是管理学中的成就与动机之间的关系。此时即会形成路径分析中所谓的非递归模型（如图 7.3b 所示）。在非递归模型中，IV 与 DV 的配对是可逆的。

可逆性的回溯因果关系是一种特殊的因果关系。假设两个变量之间互为因果，或是一群变量之间具有循环式的间接关联，则造成了两变量的直接回溯效果。一般在横断研究中，两个变量互为因果的假设不容易在理论与概念层次获得清楚的界定，因为在同一个特定时间点下，不太可能某一个自变量却又成为他所解释的因变量的因变量。因此一般社会科学中的研究均不建议采用此种模型，以免造成解释上的混淆（interpretational confounding）。读者可以试想，如果既然“互为因果”，那么同一个时间下，因会随果而“变动”，果也会随因而“变动”，这种因已非因果已非果的现象，还可称之为因果吗？

（二）延宕模型

在具有时间落差的纵贯性或重复量数（repeated measures）的研究中，因为存在有限因果性延宕（finite casual lag），两个变量互为因果的假设模型比较可能被接受。因为互为因果的两个变量，在经过一段时间的延宕后，成为彼此的果与因，两个时间点的距离不论长短，都提供了“因”、“果”两变量成为他人“果”、“因”所需的时间落差，如图 7.4 的实线所示。

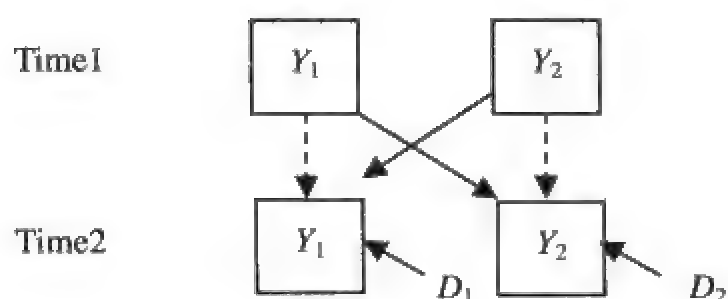


图 7.4 具有时间延宕的非递归模型

图 7.4 又称之为延宕模型 (lag model) 或跨因果模型 (cross-causal model) (Maruyama, 1997), 其重要价值在于解决了干扰具有相关的问题 (如图 7.3b 当中 D_1 与 D_2 的相关)。从图 7.4 可以看出, 除了两个实线说明了研究者所主张的互为因果的关系之外, 还存在着两个关系更强的因果关系 (由第一个 Y_1 到第二个 Y_1 、由第一个 Y_2 到第二个 Y_2), 亦即变量本身的前后期关系。但由于它并非研究者所关心, 因此以虚线表示。变量的前后期的自我相关 (auto-correlation) 如果没有纳入估计, 很显然会影响整个模型的参数估计, 而这也就是为什么图 7.3b 当中会有 D_1 与 D_2 的相关设定。延宕模式优于可逆模式的主要理由, 就是可以有效处理残差相关的问题。

Maruyama (1997) 指出, 可逆模式经常会有无法收敛的问题, 即是因为干扰项的关系太强, 参数不容易得到唯一解, 甚至因为低度识别而无解。从图 7.3b 也可以看出, 干扰项 D_1 与 D_2 如果不设定相关, 那么也存在一种可能的影响关系: $D_1 \rightarrow Y_1 \rightarrow Y_2$ 与 $D_2 \rightarrow Y_2 \rightarrow Y_1$, 造成参数估计的不稳定性 (Kaplan, 2000; Kaplan, Harik, & Hotchkiss, 2000)。相对之下, 不可逆的递归模型则容易被有效识别、获得终解, 亦具有解释明确与理论清晰的优势。

(三) 工具变量模型

Maruyama (1997) 曾介绍一种如何在非递归模型中, 藉由纳入工具变量 (instrumental variable) 协助模式得以识别的做法。亦即纳入一个变量 (工具) 解释某些内生变量, 但不影响其他内生变量协助解释内生变量间的复杂关系。

前面的讨论中已经提及, 互为因果的回溯变量关系在概念上不易解释, 在统计

上更有识别不足而无法得解的窘境。为了让模式得以识别，可以增加一些条件协助参数的有效估计。例如在模型中增加变量解释互为因果的两个变量。以图 7.5 为例，模型中外源变量共有 X_1 、 X_2 与 X_3 ，互为因果的变量仍为 Y_1 与 Y_2 ，此时共有 5 个变量，共能产生 $(5+6)/2=15$ 个数据点。模型中有 12 个待估计参数（包含实线与虚线、 D_1 与 D_2 ），因此模式得以识别。

值得注意的是，对 Y_1 而言，三个外源变量都解释 Y_1 ，因而缺乏工具变量澄清回溯关系。对 Y_2 而言， X_1 并没有解释 Y_2 ，因而可作为 Y_1 与 Y_2 的工具变量，换言之，由于 X_1 仅能解释 Y_1 ，而与 Y_2 无关，因此透过 $X_1 \rightarrow Y_1 \rightarrow Y_2$ 的关系说明 $Y_1 \rightarrow Y_2$ 的影响（ X_1 作为 Y_2 的工具）。

如果把图 7.5 当中的虚线移除后，那么就获得 $X_3 \rightarrow Y_2 \rightarrow Y_1$ 的关系说明 $Y_2 \rightarrow Y_1$ 的影响合理性（ X_3 作为 Y_1 的工具），如此一来自由度增加 1，也强化了模式的解释力，就是纳入工具变量的优点。但值得注意的是，外源变量之间相关不能太高，除了解释上的困难，也将造成共线性问题，进而使得工具变量的“工具”角色失效。详细有关工具变量模型的数学内涵请参考 Maruyama (1997)、Bollen (1989)、Rigdon (1995) 的论述。

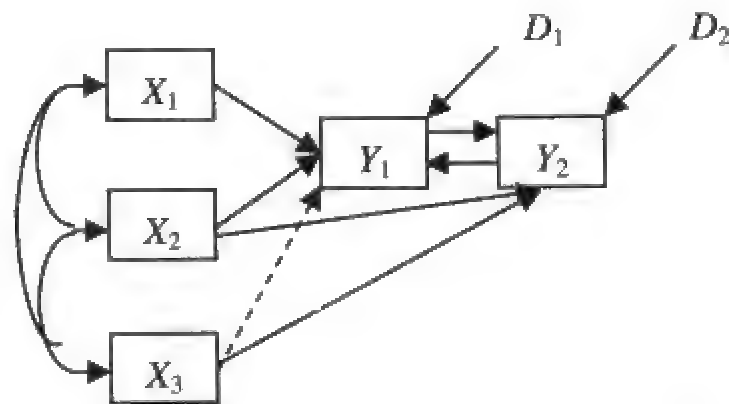


图 7.5 工具变量模型的图示

三、路径图与结构方程式

现以学生学业表现的理论模型建构为例，说明一个假设性的路径模型。如图 7.6 所示，研究者认为学生的求学动机会影响其学业表现，在参酌相关文献后，认

为学生受到他人影响的因素（社会期待）与个人对自己能力的判断（自我效能感）构成动机的基础，进而影响学业的表现。整个模型可以图 7.6 的路径图描述。图中包括了四个变量之间的结构关系，包括自我效能感、社会期待、成就动机与学业表现。单箭头代表因果方向，双箭头则代表相关。此一结构关系，系由下列三组假设所组成：

假设一：自我效能感与社会期待影响个人的成就动机

假设二：自我效能感、社会期待与成就动机影响学业表现

假设三：自我效能感与社会期待具有相关

上述假设中，假设一与假设二皆为多元回归假设，对于假设一而言，以成就动机为因变量，自我效能感与社会期待为自变量，假设二则以成就动机与自我效能感、社会期待同列为预测学业表现的自变量。可以下列方程式表现：

$$Y_1(\text{成就动机}) = b_1X_1(\text{自我效能感}) + b_2X_2(\text{社会期待}) + a_1$$

$$Y_2(\text{学业表现}) = b_3X_1(\text{自我效能感}) + b_4X_2(\text{社会期待}) + b_5X_3(\text{成就动机}) + a$$

这两个方程式构成一套结构方程式，即称为路径模型（causal model），各变量与因变量之间的关系系数 b_i 称为路径系数（path coefficient）。路径模型除了必须满足回归分析的所有假设，例如，变量之间的关系均必须是线性且具有可加性（linear and additive），变量皆属可量化的连续变量，误差项为正态且独立，等等。

图 7.6 中，成就动机同时具备自变量（ X_3 ）与因变量（ Y_2 ）的双重身份。自我效能感与社会期待仅作为自变量不受其他变量的影响，是为外源变量，其变异量由不属于路径模型的其他变量所决定。外源变量之间可能具有相关，也可能相互独立，但它们的关系并不影响路径模型内的因果关系。

图 7.6 的模型中，学业表现则纯为因变量，其变异量完全由路径模型中的其他变量的线性组合所决定，是为内生变量， D 则为无法被变量解释的独特变异。值得注意的是，成就动机变量同时具备自变量与因变量的双重身份。其自身的变异量，由路径模式中的自我效能感与社会期待两个自变量所决定，因此也属于内生变量。

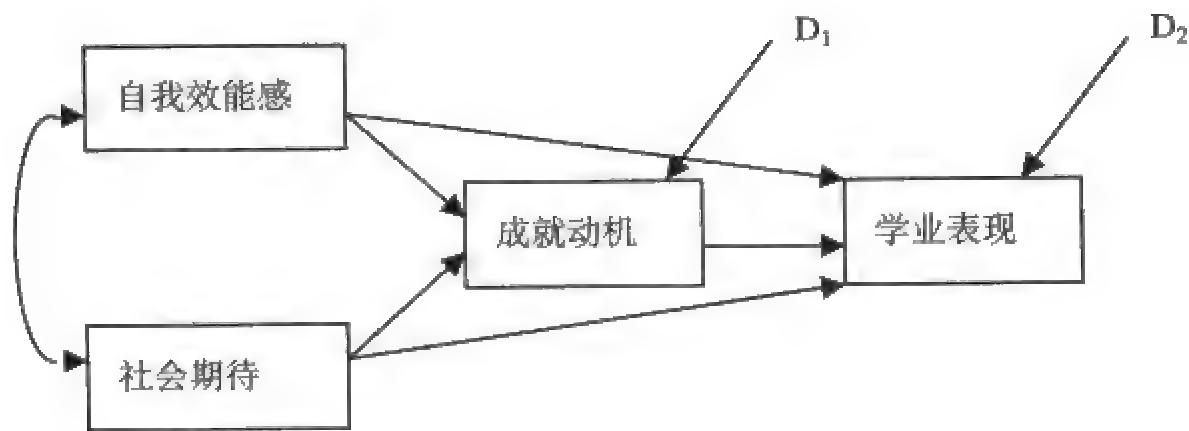


图 7.6 影响学业表现的路径模型图

四、直接效应与间接效应

经由上述的描述，变量之间的关系得以结构的方式呈现，每一个独立的回归方程式，可以回归分析的原理与技术进行分析。对于每一个回归模式，自变量对于因变量的解释力可以由 R^2 及 F 检验值表示，而每一个箭头则可以获自于回归系数，无论显著或不显著的回归系数均需填注于路径模式箭头两侧。箭头的回归系数若达显著，代表该因果变量间具有直接效应 (direct effect)，未显著的箭头回归系数，则代表无直接效应存在。两个变量之间，除了可能具有直接效应，亦可能存在间接效应 (indirect effect)，也就是说，两个变量间，具有一个或多个中介变量 (mediated variable)，变量与变量之间的直接效应均为显著，若有任何一个直接效应不显著，则间接效应无法成立。

(一) 效应分解 (effect decomposition)

图 7.7 是一个路径分析的结果。自我效能感对于成就动机与学业表现均有直接效应，标准化回归系数 β (路径系数) 分别为.29 ($p < .01$) 与.63 ($p < .001$)，同时由

于成就动机对于学业表现亦有显著的直接效应，因此，自我效能感对于学业表现的影响（ $\beta=.21, p<.01$ ），除了具有直接效应之外，尚具有一个由成就动机所中介的间接效应。

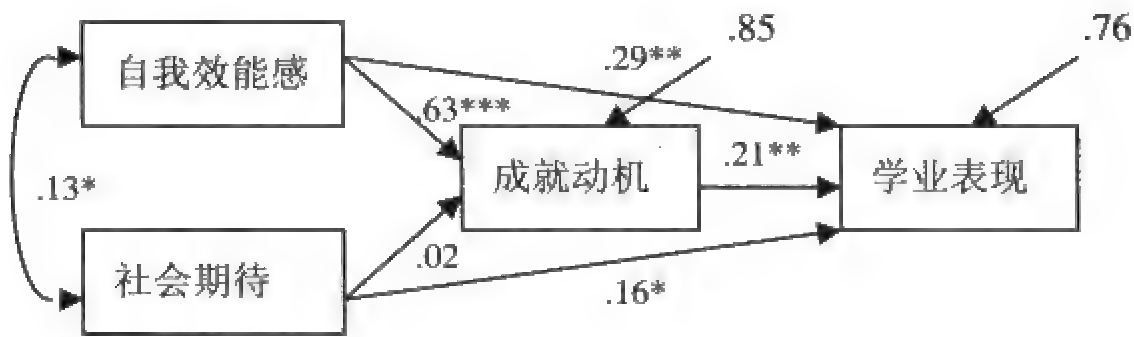


图 7.7 路径分析参数估计图

对于社会期待而言，由于对于成就动机的预测力不足（ $\beta=.02, n.s.$ ），直接效应不明确，但是对于学业表现仍具有直接预测力（ $\beta=.16, p<.05$ ），然而，社会期待已无法藉由此间接影响学业表现。

间接效应的强度可直接由两端点变量之间的直接效应标准回归系数相乘而得。自我效能感对于学业表现的间接效应由两个直接效应（自我效能感→成就动机，成就动机→学业表现）所组成，取两者的回归系数相乘得到间接效应 $.63 \times .21 = .13$ ，代表每一标准差单位的两个自变量的变动，对于学业表现造成的变动量为.13个单位。

每一个自变量对于每一个内生变量的整体效应（total effect），可以从路径模型当中与该自变量与内生变量有关的所有显著与不显著的直接效应与间接效应的回归系数值加总而得之。以自我效能对学业表现的整体效应为例，取 $.29$ （自我效能→学业表现） $+.13$ （自我效能→成就动机→学业表现） $=.42$ ，代表每一标准差单位的自变量（自我效能感）的变动，对于学业表现造成的变动量为.79个单位。

对于自我效能对成就动机的影响而言，整体效应即为唯一的一个直接效应.63，无间接效应的存在。该值表示每一标准差单位的所有自变量的变动（自我效能感）对于学业表现造成的整体变动量为.63个单位。上述所有的结果可以整理如表 7.2。

表 7.2 回归取向路径分析各项效应分解说明

自变量	内生变量	
	成就动机	学业成绩
自我效能感		
直接效应	.63***	.29**
间接效应	-	.13*
整体效应	.63***	.42**
社会期待		
直接效应	.02	.16*
间接效应	-	.00
整体效应	.02	.16*
成就动机		
直接效应		.21**
间接效应		-
整体效应		.21**

* $p < .05$; ** $p < .01$; *** $p < .001$

(二) 模型衍生相关与模型拟合

表 7.3 整理了路径模型中各变量的影响效果，传统上，以一般回归分析（OLS）即可获得这些参数值。即使利用 SEM 软件来估计，数据也会相同。

从统计的角度来看，任何两变量之间都可以计算出一个协方差反映两个变量的关系，称为观察相关（observed correlation），但表 7.3 所列出的是经由路径模型推导并估计所得到的回归效果。Pedhazur（1997）详细说明了在路径分析中，如何将间接效应、直接效应加以合并，并纳入未被分析的拟似相关（spurious correlation）（图 7.5 当中的自我效能与社会期待的相关系数），计算出路径模型中任两变量的模型衍生相关（model-implied or predicated correlation）。计算的原理称为轨迹法则（tracing rule）。

所谓的“轨迹”主要在寻找由于相关系数所导致的间接效应。对于每一个自变量对于内生变量的模型衍生相关，系由先前所求得的整体效应再加上尚未被计算的相关导致的间接效应。以图 7.7 为例，自我效能感对于学业成绩的模型衍生相关，总

计牵涉四条轨迹：

1. 直接效应：自我效能 \rightarrow 学业表现 $=.29$
2. 间接效应：自我效能 \rightarrow 成就动机 \rightarrow 学业表现 $=.13$
3. 相关间接效应（拟似效应）I：自我效能 \leftrightarrow 社会期待 \rightarrow 学业表现 $=.13 \times .16 = .02$
4. 相关间接效应（拟似效应）II：自我效能 \leftrightarrow 社会期待 \rightarrow 成就动机 \rightarrow 学业表现 $=.13 \times .02 \times .21 = .00$

因此，自我效能感对于学业成绩的模型衍生相关如下：

$$.29 + .13 + .02 + .00 = .44$$

此一相关系数可以与该两变量的简单相关（例如.46）进行比较，两者的差距代表理论模式与观察数值的差距（残差相关），模型中所有其他两两变量的模式衍生相关与简单相关的差异分别列出，可以得到整体模型与观察值的整体差异程度的残差相关矩阵。最终得以了解模式拟合情形。

（三）路径系数、中介效应与 Sobel 检验

从最简单的三变量关系来看， $X \rightarrow Me \rightarrow Y$ 的间接效应其实就是中介效应。换言之， X 对 Y 的影响力除了可以从直接效应来看，更重要的是透过 Me 的中介效应来解释，甚至只有当 $X \rightarrow Y$ 的效应从原来不考虑 Me 的情况下为显著，而考虑 Me 之后变成不显著时，才被称为完全中介效应（两个变量间只有间接效应而无直接效应）（Baron & Kenny, 1986）。显然，间接效应可以说是路径分析最重要的焦点所在。

间接效应的强度简单来说就是取 $X \rightarrow Me \rightarrow Y$ 的两个标准化回归系数相乘，又称为路径系数（path coefficient）（Wright, 1960）或结构系数（structural coefficient）（Duncan, 1975）。间接效应的强度是否具有统计的显著性，可以利用前述介绍过的 Sobel（1982）所推导的标准误来计算 t 检验值：

$$t = \frac{\beta_a \beta_b}{\sqrt{S_{\hat{\beta}_b}^2 \hat{\beta}_a^2 + S_{\hat{\beta}_a}^2 \hat{\beta}_b^2}} \quad (7.8)$$

上式分母中的数值可以利用传统 OLS 回归分析或 SEM 的估计数中获得。但由于两个正态化的回归系数相乘后并不服从正态分布（呈现峰度为 6 的高狭峰分布）（Lomnicki, 1967; Springer & Thompson, 1966），如果变量的平均数不为零时，还有偏态问题，因此采用 Sobel (1982) 的公式 (7.8) 导出的标准误为偏估计值 (biased estimator)。Sampson 与 Breuning (1971) 导出了路径系数的不偏估计数，得以进行路径系数的显著性检验：

$$t^* = \frac{\beta_a \beta_b}{\sqrt{S_{\hat{\beta}_b}^2 \hat{\beta}_a^2 + S_{\hat{\beta}_a}^2 \hat{\beta}_b^2 - S_{\hat{\beta}_b}^2 S_{\hat{\beta}_a}^2}} \quad (7.9)$$

虽然公式 (7.9) 修正了非正态问题，但是当样本数太小时（少于 200）常会发生估计数为负值的非正定问题而无法有效估计 (Mackinnon et al., 2002; Mackinnon, Lockwood, & Williams, 2004)。此外，Bobko 与 Rieck (1980) 建议在进行路径分析前，先将变量进行标准化，并利用 X 、 Me 、 Y 三者的相关系数计算标准误。虽然标准误的公式相继被提出，但经过模拟研究发现，Sobel (1982) 所提出的公式 (7.8) 仍是效率最佳的路径系数标准误 (Mackinnon, 2008)，这也是为何多数的 SEM 软件（例如 LISREL、EQS、MPLUS）仍以 Sobel (1982) 作为间接效应的显著性检验方法。

除了显著性检验之外，中介效应的强度也可以利用区间估计来描述，并得以进行不同的中介效应的相互比较，公式如下：

$$95\%CI = \hat{\beta}_a \hat{\beta}_b \pm 1.96 S_{\hat{\beta}_a \hat{\beta}_b} = \hat{\beta}_a \hat{\beta}_b \pm 1.96 \sqrt{S_{\hat{\beta}_b}^2 \hat{\beta}_a^2 + S_{\hat{\beta}_a}^2 \hat{\beta}_b^2} \quad (7.10)$$

公式 (7.10) 是 Sobel (1982) 的路径系数标准误，如果要考虑非路径系数的正态问题，可以利用非对称分布区间估计表示置信区间。此时公式 (7.10) 当中的临界值就不再是 1.96，而分别是两个数值（有关非对称分布的计算程序、统计原理请参考 Mackinnon (2008)、Mackinnon, Fritz, Williams, & Lockwood (2007)，相关信息与软件 RIPL 可在 <http://www.public.asu.edu/~davidpm/ripl/index.htm> 获得）。

(四) 多中介变量的效应检验

前面所讨论的是当模型中只有一个中介变量的情况，但研究中可能同时存在多个中介变量，因此路径系数的检验与比较就开始变得复杂起来。图 7.8 显示了当 $X \rightarrow Y$ 之间存在两个中介变量 M_1 与 M_2 时的情形。

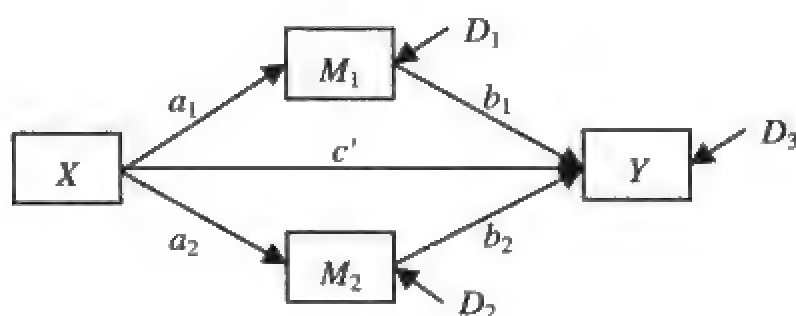


图 7.8 带有两个中介变量的路径模型

图 7.8 当中的路径系数有二： $X \rightarrow M_1 \rightarrow Y$ 的 a_1b_1 与 $X \rightarrow M_2 \rightarrow Y$ 的 a_2b_2 ，这两个中介作用的总和，就是 $X \rightarrow Y$ 原来的简单回归系数降至路径模型当中 c' 的减少量：

$$c - c' = a_1b_1 + a_2b_2 \quad (7.11)$$

利用传统的 OLS 多元回归可以得到所有系数的估计值与标准误（以 SEM 软件所估计的数值与 OLS 估计数相近）。值得注意的是，如果图 7.8 的模型当中各方程式的样本数不同时，公式（7.11）的等号可能会不成立，以 SEM 软件估计时，由于采用同一个协方差矩阵，整个模型只有一个样本数，这种情形就不至于发生。

图 7.8 当中， a_1b_1 与 a_2b_2 两个各别的中介作用显著性可以由先前的公式（7.8）来检验，但是由 $X \rightarrow Y$ 的总间接效应检验，则必须将两者效应加总后，计算合成标准误来进行 t 检验，标准误公式如下：

$$S_{\hat{a}_1\hat{b}_1 + \hat{a}_2\hat{b}_2} = \sqrt{S_{\hat{a}_1}^2\hat{b}_1^2 + S_{\hat{b}_1}^2\hat{a}_1^2 + S_{\hat{a}_2}^2\hat{b}_2^2 + S_{\hat{b}_2}^2\hat{a}_2^2 + 2\hat{a}_1\hat{a}_2S_{\hat{b}_1\hat{b}_2}} = \sqrt{S_{\hat{a}_1\hat{b}_1}^2 + S_{\hat{a}_2\hat{b}_2}^2 + 2\hat{a}_1\hat{a}_2S_{\hat{b}_1\hat{b}_2}} \quad (7.12)$$

如果要比较的是两个中介效应的差异是否具有统计的显著性，可以下式来检定：

$$S_{\hat{a}_1\hat{b}_1-\hat{a}_2\hat{b}_2} = \sqrt{S_{\hat{a}_1\hat{b}_1}^2 + S_{\hat{a}_2\hat{b}_2}^2 - 2\hat{a}_1\hat{a}_2S_{\hat{b}_1\hat{b}_2}} \quad (7.13)$$

公式 7.12 与 7.13 中， $S_{\hat{b}_1\hat{b}_2}$ 是 \hat{b}_1 与 \hat{b}_2 两个回归系数的协方差。但如果 \hat{a}_1 与 \hat{a}_2 两个回归系数的协方差不为 0，则两式的根号当中还要增加一项 $2\hat{b}_1\hat{b}_2S_{\hat{a}_1\hat{a}_2}$ 。

当路径模型更趋于复杂时，例如 $X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$ ，或有更多的中介变量时，路径系数与标准误的计算原理类似，只是所需考虑的参数数目更多，超过了人工计算的能力。因此在此不予介绍计算式内容。此时最好的处理方式是利用 SEM 软件以单一的协方差矩阵来同时估计各参数与标准误 (Sobel, 1986, 1987)。这些软件同时也会计算出间接效应（某特定 $X \rightarrow Y$ 关系中的中介效应的总和）的强度与显著性检验结果。

五、结构方程模型的路径分析

先前提及第三变量将使 $X \rightarrow Y$ 的关系产生相当复杂的变量，如果一个模型存在第四或第五个变量，那么整个模型就相当复杂。而路径分析就是一种利用回归模型整合一连串复杂的变量关系所形成的完整模型。模型中所有变量的关系虽然未必全然是中介关系，也可能带有调节变量与非线性关系，但在本章的架构下，主要系以中介模型来分析变量关系。

其次，一般当研究者提及路径分析一词时，主要是指一组外显观察变量的关系，而不涉及潜在变量的定义与分析。近年来，由于 SEM 盛行，因此路径分析一词也逐渐与结构方程模型当中的结构模型分析混淆并用。一般而言，当一个模型称之为结构方程模型，是指模型中除了带有诸多变量间的因果解释关系之外，各研究变量还牵涉到潜在变量的定义，因而有学者将之称为带有潜在变量的结构方程模型 (structural equation modeling with latent variables) (Tabachnicka & Fidell, 2007)。至于路径分析则是指仅包含外显变量的多重因变量的复杂回归模型，不涉及潜在变量的定义。近来由于结构方程模型的软件逐渐普及，以及以 SEM 软件进行路径分析有诸多优点，因此绝大多数的路径分析已采用 SEM 软件以完整的共变结构进行分析。为了有别于带有潜在变量的路径分析一词，Tabachnicka 与 Fidell (2007) 将其称之为

带有观察变量的结构方程模型 (structural equation modeling with observed variables)。本章首先介绍路径分析的操作与解释, 至于潜在变量模型则待第九章介绍之。

以结构方程模型进行路径分析, 可以说是 SEM 最重要的应用价值之一。如果单从取代性来看, SEM 的分析软件与技术原理可以完全取代过去路径分析的分析任务, 更可以超越过去路径分析只能以外显变量作为分析变量的限制, 透过统合模型分析 (hybrid modeling) 的运用, 巧妙地将因素分析的概念与技术融合到路径分析的检验中, 堪称是计量技术的一大革命。

传统的路径分析中, 所处理的变量若以 SEM 术语表示, 都应该以方块表示其性质属于测量变量, 而无椭圆形的潜在变量。亦即路径分析当中的变量是假设没有测量误差, 即使误差明显 (例如信度很低); 在路径分析过程中, 变量的误差也将被忽略。SEM 取向的路径分析可以说是 SEM 的一种应用特例, 也就是没有包含任何潜在变量的结构模型分析。在 SEM 路径图中, 变量皆以方块表示。结构方程式通式如下:

$$Y = \alpha + BY + \Gamma X + \zeta \quad (7.14)$$

其中 B 为内生变量之间的回归系数 (Beta) 矩阵, Γ 为外源变量与内生变量间的回归系数 (Gamma) 矩阵, α 为截距, ζ 为回归残差 (disturbance)。当 SEM 分析共变或相关矩阵的关系时, α 为 0, 因此被忽略。此外, 外源变量间的相关并没有反映在上述通式当中, 但会在 Φ 矩阵中加以估计。

以图 7.8 的双中介变量模型为例, SEM 取向的共变结构分析的参数如图 7.9 所示。以结构方程式表示如下:

$$Y_1 = \gamma_{11}X_1 + \zeta_{11} \quad (7.15)$$

$$Y_2 = \gamma_{21}X_1 + \zeta_{22} \quad (7.16)$$

$$Y_3 = \gamma_{31}X_1 + \beta_{31}Y_1 + \beta_{32}Y_2 + \zeta_{33} \quad (7.17)$$

以矩阵表示如下:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \end{bmatrix} [X_1] + \begin{bmatrix} \zeta_{11} \\ \zeta_{22} \\ \zeta_{33} \end{bmatrix}$$

由于两个中介变量 (Y_1 与 Y_2) 之间可能存在相关, 因此图 7.9 中的解释残差 ζ_{11} 与 ζ_{22} 之间可以设定一个自由估计的参数 ζ_{12} (虚线所示)。此时残差变异矩阵 (Ψ) 与外源变量方差矩阵 (Φ) 如下:

$$\Psi = \begin{bmatrix} \psi_{11} & & \\ \psi_{21} & \psi_{22} & \\ 0 & 0 & \psi_{33} \end{bmatrix}$$

$$\Phi = [\phi_{11}]$$

此模型的待估参数数目为 10, 数据点为 $(4+5)/2=10$, 两者相同, 为一恰为识别 (just-identified) 的饱和模型 (saturated model), $df=0$, $\chi^2=0$, 参数有唯一解, 估计值的数据与传统 OLS 回归分析的结果相近。由于模型拟合指数为 0, 因此无法进行模型优劣的检验。除非移除模型中的某些待估参数, 例如 ζ_{12} 或 γ_{31} 。此时路径模式背后的理论观点就略有不同。

如果图 7.9 再增加任何一个参数 (例如 ζ_{23}), 将造成识别不足 (under-identified) 的问题, 模型无法收敛估计。换言之, 只有当模型自由度大于 0 时的非饱和模式 (non-saturated model), 拟合指数才得以计算。当模型为可识别的情况下 (自由度大于 0), 研究者可采模式竞争策略, 藉由增减参数比较不同的嵌套模型 (nested model) (参数少者的较小模型嵌套在参数较多的较大模型当中), 此时模型的优劣可以利用两个模型的卡方差异量 ($\Delta\chi^2$) 是否达到自由度为两个模型自由度差异量 (Δdf) 下的统计显著水平检验两者拟合是有否有别 (亦即额外增加的参数是否能够改善模式拟合性), 称为卡方差异检验 (chi-square difference test)。有关模式竞争比较的具体内容与统计基础可参见 Bollen (1989)、Kaplan (2000)、Kline (1998) 的讨论。

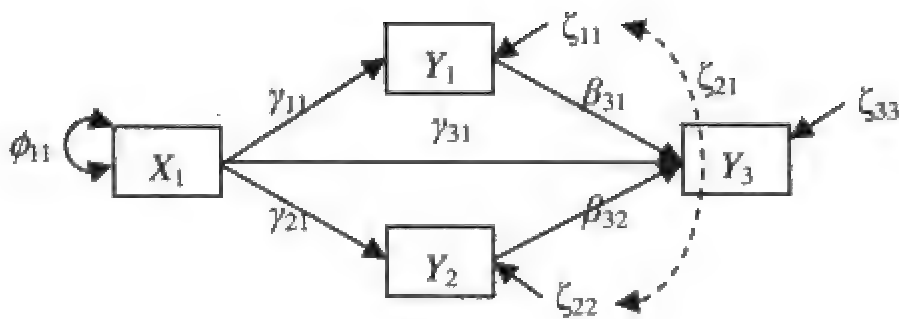


图 7.9 双中介变量的路径模型共变结构分析图

第三节 LISREL 的路径分析

为了说明路径分析的操作，我们将以某上市公司 289 位员工的组织行为研究调查数据作为范例。数据库中的测量变量包括组织气氛的知觉、组织承诺与员工绩效。为了符合路径分析的形式，本范例将量表的得分加总得到一个总分，以外显观察变量的形态进行分析。组织气氛的测量曾在第五章的范例中详细介绍，共有六个方面，包括“组织价值”、“工作方式”、“团队合作”、“领导风格”、“学习成长”、“环境气氛”。员工的组织承诺则是将组织承诺量表的得分加总后，得到一个单一的承诺分数，分数越高，代表员工的向心力与承诺意愿高。最后，员工绩效变量则由三个题目的平均数表示，这三个题目分别为工作满意度、离职意愿、自我绩效评定，以 1 至 10 分的量尺测得，员工绩效变量分数越高，显示员工工作状况良好。另外加上年资变量（作为控制变量），总共有 9 个观察变量。以下，我们将利用 SEM 的结构模式分析进行路径分析的估计程序。

一、模型界定

本研究的主要假设是组织气氛的知觉影响员工的承诺感，进而影响员工的工作表现。此时组织承诺扮演中介变量的角色，年资变量则是一个干扰变量或控制变量，其角色不但影响了组织承诺，也影响工作产出，它调节了承诺感与员工绩效的预测力。路径图如图 7.10 所示。

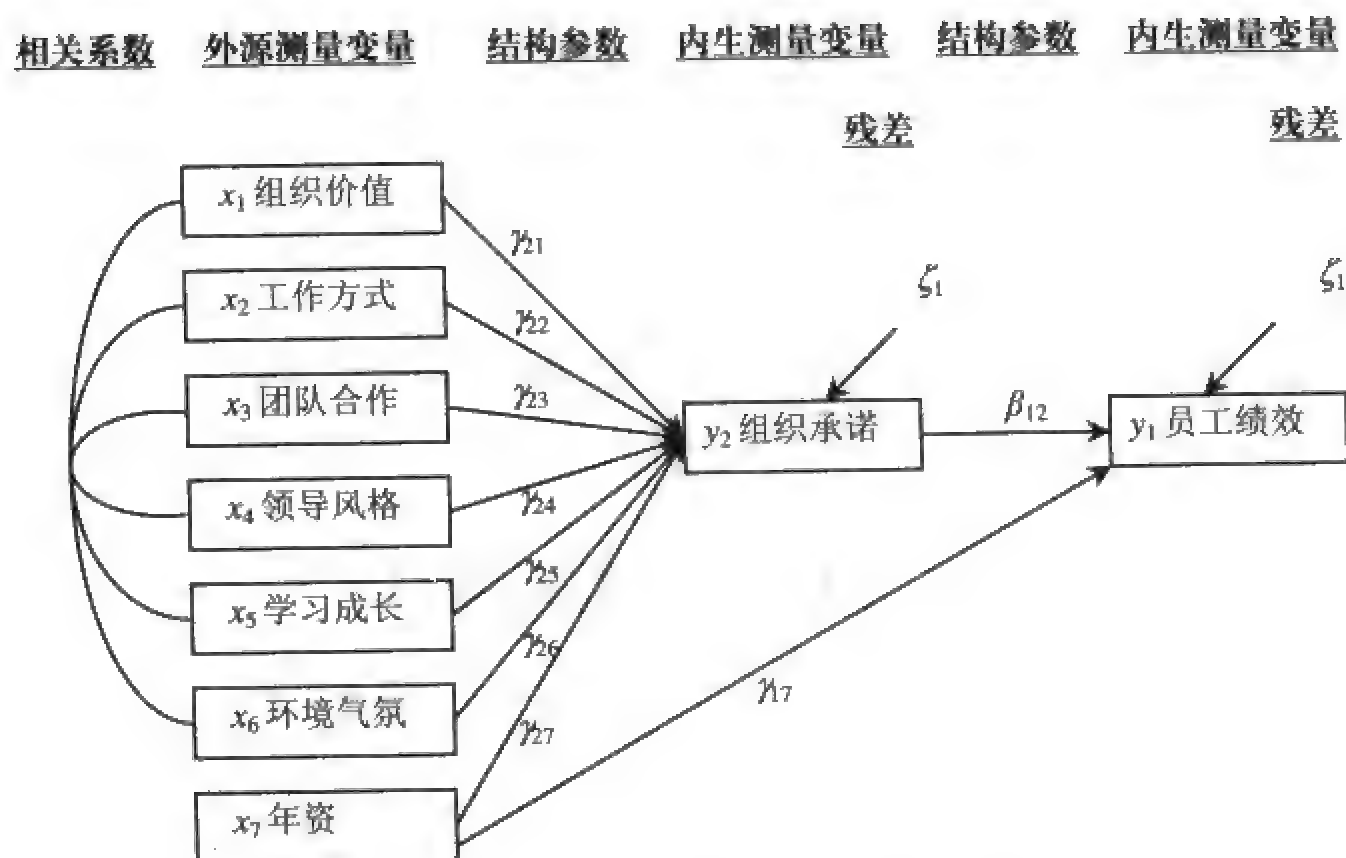


图 7.10 路径分析的假设模型路径图

图 7.10 说明了研究变量的假设关系，也列出了路径分析的主要参数。当以组织承诺为因变量时 (Y_2)，六个组织气氛变量 (X_1 至 X_6) 与年资变量 (X_7) 为自变量，当以员工绩效为因变量 (Y_1) 时，仅有年资与组织承诺为自变量，整个模型的回归方程式如下：

$$Y_1 = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + a_1$$

$$Y_2 = b_7 X_7 + b_8 X_8 + a_2$$

方程式中, b 与 a 分别为回归系数(斜率)与截距。SEM 的路径分析中, 截距设定为 0, 回归系数视其性质则属于不同的参数矩阵。

本范例中,整个模型的观察变量共有 9 个,因此测量数据数为 $(9 \times 10)/2 = 45$ (DP=45)。七个外源变量之间都可能具有相关,但是图 7.10 显示年资变量与其他六个组织气氛的

测量并未假设具有相关,因此在模型界定时,需予以说明。

模型中有 0 个测量残差,但两个内生变量有 2 个解释残差 (ζ_1 、 ζ_2),其变异量被自由估计。第一个内生变量(员工绩效)被两个外源变量解释,故 BETA 矩阵中有 2 个结构参数 (β_1 、 β_2)。第二个内生变量(组织承诺)被七个外源变量解释,故 GAMMA 阵中有 7 个结构参数 (γ_1 至 γ_7)。6 个外源变量之间的相关允许自由估计,表示 ϕ 矩阵将产生 30 个相关系数。

二、LISREL 语法

本范例所输入的数据为方差与协方差矩阵,9 个测量变量总共产生了 9 个方差与 36 个协方差,以对称三角形的矩阵方式列出。值得注意的是,LISREL 在输入变量时,先读入内生变量再读入外源变量,因此,本范例输入的变量顺序分别如下:员工绩效(y1)、组织承诺(y2)、组织价值(x1)、工作方式(x2)、团队合作(x3)、领导风格(x4)、学习成长(x5)、环境气氛(x6)、以及年资(x7)。依序读入各变量的方差与协方差。LISREL 语法如表 7.3 所示。

表 7.3 的 LISREL 语法,共有 9 个主要的指令。DATA 指令说明变量数目为 9,样本数目 281,数据以协方差矩阵形式读入(MA=CM)。自第 3 个指令开始,CM SY 指出下列为对称协方差矩阵,依序读入所有的变量的方差与协方差。并在第 4 个指令中予以卷标。

第 5 个指令指出外源观察变量数目为 2 (NY=2),内生观察变量数目为 7 (NX=7),内生变量残差矩阵为对角线矩阵并自由估计 (PS=DI,FR),相关系数矩阵为对称矩阵并估计之 (PH=SY,FR),结构系数矩阵则为完全矩阵并先加以固定 (BE=FU,FI GA=FU,FI),所需估计的结构参数则由 FREE 指令指定之。分别为第 6 个指令中的 9 个参数。最后,第 7 个指令逐一列出不予以估计的 6 个相关系数,指出除了与年资有关的相关系数不估计之外,其余的相关系数均予以估计。输出指令中比较重要的是 EF 指令,可要求 LISREL 列出总效应与间接效应。

表 7.3 路径分析的 LISREL 语法范例(ch7a.ls8)

1	Ch7a Path Modela using LISREL syntax
2	DATA NI=9 NO=281 MA=CM
3	CM SY .402 .466 2.537 .185 .667 .649 .174 .704 .368 .910 .167 .578 .272 .421 .693 .157 .483 .207 .330 .320 .600 .221 .730 .346 .463 .392 .387 .779 .172 .647 .370 .270 .369 .250 .333 .768 1.406 2.883 .949 .154 .347 .274 .527 1.116 85.137
4	LA; OUTCOME COMMIT VALUE JOBSTYLE TEAMWORK LEADERSH LEARNING ENVIRONM TENURE
5	MODEL NY=2 NX=7 PS=DI,FR PH=SY,FR BE=FU,FI GA=FU,FI
6	FREE BE(1,2) GA(2,1) GA(2,2) GA(2,3) GA(2,4) GA(2,5) GA(2,6) GA(2,7) GA(1,7)
7	FIX PH(1,7) PH(2,7) PH(3,7) PH(4,7) PH(5,7) PH(6,7)
8	PD
9	OUTPUT SE TV RS MR FS EF SS SC MI

三、SIMPLIS 语法

SIMPLIS 的语法列于表 7.4, 路径模型的设定主要在第 5 至第 8 行当中, 被解释的因变量置于等号左侧, 解释变量则放于右侧。语法中没有出现 LISREL OUTPUT 的指令, 表示将列出简式的 SIMPLIS 报表。

表 7.4 路径分析的 SIMPLIS 语法范例(ch7a.spl)

```

1  Ch7a Path Modela using SIMPLIS syntax
2  Observed variables: OUTCOME COMMIT VALUE JOBSTYLE TEAMWORK LEADERSH
3  LEARNING ENVIRONM TENURE
3  Covariance matrix:
    .402
    ... (略)
4  Sample Size = 281
5  Relationships:
6      OUTCOME = COMMIT
7      OUTCOME = TENURE
8      COMMIT = VALUE JOBSTYLE TEAMWORK LEADERSH LEARNING ENVIRONM TENURE
9  Path Diagram
10 Print Residuals
11 End of Problem

```

四、结果报告与说明

本范例直接读入协方差矩阵数据, LISREL 省略了处理协方差计算的程序, 所得到的读入信息显示, 内生变量 (Number of Y Variables) 个数为 2, 外源变量 (Number of X Variables) 个数为 7。值得注意的是, 虽然潜在变量数目出现了 2 (ETA) 与 7 (KSI), 但是本范例并未针对观察变量的误差项进行估计, 因此潜在变量并不会出现在报表当中。这些摘要信息与协方差矩阵的报表如下 (参见结果档 ch7a.out):

Ch7a Path Modela using LISREL syntax

```

Number of Input Variables  9
Number of Y - Variables    2
Number of X - Variables    7
Number of ETA - Variables  2
Number of KSI - Variables  7
Number of Observations    281

```

Covariance Matrix						
	员工绩效	组织承诺	组织价值	工作方式	团队合作	领导风格
员工绩效	0.44					
组织承诺	0.45	2.63				
组织价值	0.23	0.71	0.73			
工作方式	0.17	0.72	0.34	0.97		
团队合作	0.15	0.57	0.28	0.45	0.69	
领导风格	0.18	0.52	0.27	0.30	0.29	0.66
学习成长	0.21	0.72	0.34	0.41	0.37	0.38
环境气氛	0.19	0.65	0.40	0.26	0.34	0.24
年资	1.69	3.21	1.35	0.31	0.46	0.54

Covariance Matrix			
	学习成长	环境气氛	年资
学习成长	0.78		
环境气氛	0.30	0.78	
年资	0.47	1.54	80.73

(一) 参数估计结果

接下来, LISREL 将以矩阵形态列出所有参数的估计状况, 包括被自由估计以及被固定为 0 或 1 的参数。被估计的参数依序编号列出, 依序为编号 1 的 BETA (外源变量对内生变量解释的结构系数)、编号 2 至 9 的 GAMMA (内生变量间的结构系数)、10 至 31 的 PHI (外源变量间的相关系数)、32 至 33 的 PSI (内生变量误差方差), 表示有 33 个参数被估计。总测量单位数为 45, 因此自由度为 12, 模型属于识别剩余情形, 可顺利估计。

在这些参数当中, Beta 与 Gamma 等结构参数为路径分析关心的路径参数 (path coefficients) (注 1), 也就是传统利用回归分析所得到的回归系数。

Parameter Specifications

BETA		
	OUTCOME	COMMIT
OUTCOME	0	1
COMMIT	0	0
GAMMA		

注 1: Beta 与 Gamma 等结构参数为路径分析关心的路径参数。

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
OUTCOME	0	0	0	0	0	0
COMMIT	3	4	5	6	7	8
GAMMA						
	TENURE					
OUTCOME	2					
COMMIT	9					
PHI						
	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
VALUE	10					
JOBSTYLE	11	12				
TEAMWORK	13	14	15			
LEADERSH	16	17	18	19		
LEARNING	20	21	22	23	24	
ENVIRONM	25	26	27	28	29	30
TENURE	0	0	0	0	0	0

LISREL 定义完成后, 即进行参数估计。由数据可知, LISREL 仅耗费 5 次的迭代即完成了最大似法估计。由报表可知, 第 5 个与第 6 个结构参数的 t 检验值未达 1.96 的门槛 (注 2), 表示该两个参数 (团队合作与领导风格对于组织承诺的解释) 的估计不具有统计意义。

Number of Iterations = 5

LISREL Estimates (Maximum Likelihood)

BETA		
	OUTCOME	COMMIT
OUTCOME	- -	0.17 (0.02) 7.99
COMMIT	- -	- -

GAMMA						
	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
OUTCOME	- -	- -	- -	- -	- -	- -
COMMIT	0.43 (0.12) 3.63	0.24 (0.10) 2.37	0.11 (0.12) 0.88	0.11 (0.12) 0.90	0.36 (0.12) 3.11	0.28 (0.11) 2.56

GAMMA	
	TENURE
OUTCOME	0.01 (0.00) 2.89
COMMIT	0.02 (0.01) 2.74

注 2: 由 t 值 (.88 与 .90) 可知该两参数的估计未达显著。其余所有的 Beta 与 Gamma 参数均达 .05 显著水平。

PHI 矩阵的估计结果完全达到显著水平, 表示各外源变量的方差 (对角线的数据) (注 3) 与协方差 (亦即相关系数) 具有统计上的意义。年资与其他各外源变量的关系不予以估计, 因此缺乏显著性检验 (注 4), 仅针对年资本身的方差进行检验。

PHI						
	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
VALUE	0.65 (0.05) 11.88					
JOBSTYLE	0.37 (0.05) 7.19	0.91 (0.08) 11.69				
TEAMWORK	0.27 (0.04) 6.27	0.42 (0.05) 7.75	0.69 (0.06) 11.71			
LEADERSH	0.21 (0.04) 5.25	0.33 (0.05) 6.74	0.32 (0.04) 7.36	0.60 (0.05) 11.70		
LEARNING	0.35 (0.05) 7.31	0.46 (0.06) 7.98	0.39 (0.05) 7.81	0.39 (0.05) 8.17	0.78 (0.07) 11.73	

注 3: 对角线的参数表示各外源变量的方差估计数, 检验结果均达显著水平。

ENVIRONM	0.37 (0.05) 7.83	0.27 (0.05) 5.13	0.37 (0.05) 7.54	0.25 (0.04) 5.77	0.33 (0.05) 6.62	0.77 (0.06) 11.91
TENURE	0.95	0.15	0.35	0.27	0.53	1.12
PHI						
	TENURE					
TENURE	85.14 (7.09) 12.01					

注 4：年资与其他外源变量的相关系数没有进行估计，故无显著性检验。

PSI
Note: This matrix is diagonal.

OUTCOME	COMMIT
0.31 (0.03) 11.68	1.46 (0.13) 11.68

Squared Multiple Correlations for Structural Equations

OUTCOME	COMMIT
0.24	0.42

注 5：表示内生变量的方差被解释的程度，类似于 R^2 的功能。

PSI 矩阵的估计结果，显示两个内生变量被解释之后的测量残差达到显著水平。而两个变量被解释的比例，分别为员工绩效的.24 与组织承诺的.42，此两个数值的意义类似于回归分析的 R^2 （注 5），只有回归分析的两个 R^2 是独立估计的结果，而此地的 Squared Multiple Correlations for Structural Equations 则为两个方程式同时估计的结果，严谨度较高。

（二）模型拟合度分析

由数据可知，整体而言，本模型适合度堪称理想，WLS 卡方值 18.59 ($p=.099$)，NNFI、CFI、GFI 均达.95 标准，然而从 RMSEA 系数 (.045) 可知，本范例所提出的假设模型良好。

Degrees of Freedom = 12
 Minimum Fit Function Chi-Square = 19.24 (P = 0.083)
 Normal Theory Weighted Least Squares Chi-Square = 18.59 (P = 0.099)
 Estimated Non-centrality Parameter (NCP) = 6.59
 90 Percent Confidence Interval for NCP = (0.0 ; 22.39)

 Minimum Fit Function Value = 0.069
 Population Discrepancy Function Value (F0) = 0.024
 90 Percent Confidence Interval for F0 = (0.0 ; 0.082)
 Root Mean Square Error of Approximation (RMSEA) = 0.045
 90 Percent Confidence Interval for RMSEA = (0.0 ; 0.083)
 P-Value for Test of Close Fit (RMSEA < 0.05) = 0.54

 Chi-Square for Independence Model with 36 Degrees of Freedom = 1585.88
 Independence AIC = 1603.88
 Model AIC = 84.59
 Saturated AIC = 90.00
 Independence CAIC = 1645.63
 Model CAIC = 237.66
 Saturated CAIC = 298.73

 Normed Fit Index (NFI) = 0.99
 Non-Normed Fit Index (NNFI) = 0.99
 Parsimony Normed Fit Index (PNFI) = 0.33
 Comparative Fit Index (CFI) = 1.00
 Incremental Fit Index (IFI) = 1.00
 Relative Fit Index (RFI) = 0.96

 Critical N (CN) = 382.62

 Root Mean Square Residual (RMR) = 0.024
 Standardized RMR = 0.045
 Goodness of Fit Index (GFI) = 0.99
 Adjusted Goodness of Fit Index (AGFI) = 0.95
 Parsimony Goodness of Fit Index (PGFI) = 0.26

(三) SIMPLIS 报表

前以 SIMPLIS 所得到的报表是以方程式的方式来表现各参数估计的结果（包括原始估计数、标准误、 t 检验值、残差方差与检验值），方程式的最后列出了各方程式的解释变异量（ R^2 ），分别为 Outcome 的.24 与 Commit 的.42。只是因为系统问题而产生乱码（Rý）。

Ch7a Path Modela using SIMPLIS syntax

Number of Iterations = 5

LISREL Estimates (Maximum Likelihood)

Structural Equations

OUTCOME = 0.17*COMMIT + 0.011*TENURE, Errorvar.= 0.31 , R² = 0.24
 (0.021) (0.0037) (0.026)
 7.99 2.89 11.68

COMMIT = 0.43*VALUE + 0.24*JOBSTYLE + 0.11*TEAMWORK + 0.11*LEADERSH +
 (0.12) (0.10) (0.12) (0.12)
 3.63 2.37 0.88 0.90

0.36*LEARNING + 0.28*ENVIRONM + 0.022*TENURE,
 (0.12) (0.11) (0.0080)
 3.11 2.56 2.74

Errorvar.= 1.46 , R² = 0.42
 (0.13)
 11.68

五、分析结果与参数报告

利用 SEM 进行路径分析可以获得相当丰富的信息。除了传统的参数估计可以一次完成之外, SEM 还提供了修正的建议, 包括了残差项之间的估计, 显示 SEM 对于传统路径分析是一个非常便利的分析工具。SEM 的另一项更有价值的应用是在最后的报表中。SEM 会直接列出路径分析所可能产生的各种直接与间接效应, 使得研究者可以立即获得路径参数的综效, 来进行研究数据的分析报告。以下, 我们就 SEM 所产生的各种最终解答进行说明。

(一) 路径参数

以 SEM 进行路径分析所产生的数据中, 以 BETA 与 GAMMA 两种结构参数最为重要, 也就是传统路径分析的路径系数。这些数据可以从标准化解 (Standardized Solution) 中获得, 数据列于下方。值得注意的是, 先前已经证实, 团队合作与领导风格无法有效解释组织承诺, 因此, 这两个参数的数值是不显著的 (注 6), 其余各参数均达显著水平。在绘制路径图时, 该两参数需标示为不显著。

Standardized Solution

BETA							
	OUTCOME	COMMIT					
OUTCOME	---	0.43					
COMMIT	---	---					
GAMMA							
	VALUE	JOBSTYLE	TEAMWORK	LEADERSHIP	LEARNING	ENVIRONMENT	
OUTCOME	---	---	---	---	---	---	
COMMIT	0.22	0.14	0.06	0.05	0.20	0.15	
GAMMA							
	TENURE						
OUTCOME	---						
COMMIT	0.16						
	0.13						

注 6: 参数估计并未达显著, 因此不具有统计意义。

注 6: 参数估计显著性未达显著, 因此数值不具有统计意义。

(二) 外源变量相关

以 SEM 进行路径分析的一个优点, 是可以整合所有参数的估计。过去在路径分析估计外源变量的相关时, 必须在回归分析之外独立计算, 增加了统计决策错误的可能性。SEM 则可以在估计结构参数之时, 一并获得外源变量间的相关系数。分析结果列于下方, 先前的显著性检验已经证实, 这些相关系数均达显著水平。

值得注意的是, 年资变量与其他六个外源变量的相关也在报表中列出 (注 7), 其数值大小显示可能达到显著水平, 但是我们并未将这六个相关系数纳入模型中, 因此, 即使这些系数有统计意义, 仍不应报告于最后的结论当中, 除非研究者修改其假设模型。

Correlation Matrix of Y and X

PSI

	OUTCOME	COMMIT	VALUE	JOBSTYLE	TEAMWORK	LEADERSH
OUTCOME	1.00					
COMMIT	0.46	1.00				
VALUE	0.24	0.52	1.00			
JOBSTYLE	0.20	0.46	0.48	1.00		
TEAMWORK	0.19	0.44	0.41	0.53	1.00	
LEADERSH	0.17	0.39	0.33	0.45	0.50	1.00
LEARNING	0.23	0.52	0.49	0.55	0.53	0.57
ENVIRONM	0.22	0.46	0.52	0.32	0.51	0.37
TENURE	0.24	0.20	0.13	0.02	0.05	0.04

Correlation Matrix of Y and X

	LEARNING	ENVIRONM	TENURE
LEARNING	1.00		
ENVIRONM	0.43	1.00	
TENURE	0.06	0.14	1.00

PSI

Note: This matrix is diagonal.

OUTCOME	COMMIT
0.76	0.58

注 7: 年资与其他变量的相关并不属于被估计的参数, 分析结论不应报告这些数据。

上述这些资料汇整之后, 研究者即可以整理出一个最终的估计后路径图, 如图 7.11 所示。图中清楚地说明了各变量的关系与参数数值大小。

值得注意的是, 图 7.11 系由 LISREL 软件直接产生, 使用者可从自报表中获得, 唯一的缺点是 LISREL 并不会加注这些参数的显著性, 甚至于不显著的参数也会列出。同时, 外源变量的相关系数以及类似于 R^2 的系数值也不会随图列出, 必须以人工方式填入图中。建议使用者可以自行绘制路径图说明最后的研究结果。

(三) 回归系数矩阵

除了前述的 Beta 与 Gamma 结构参数的数据, LISREL 另外报告了所有外源变量对于内生变量的回归系数, 列于下方。

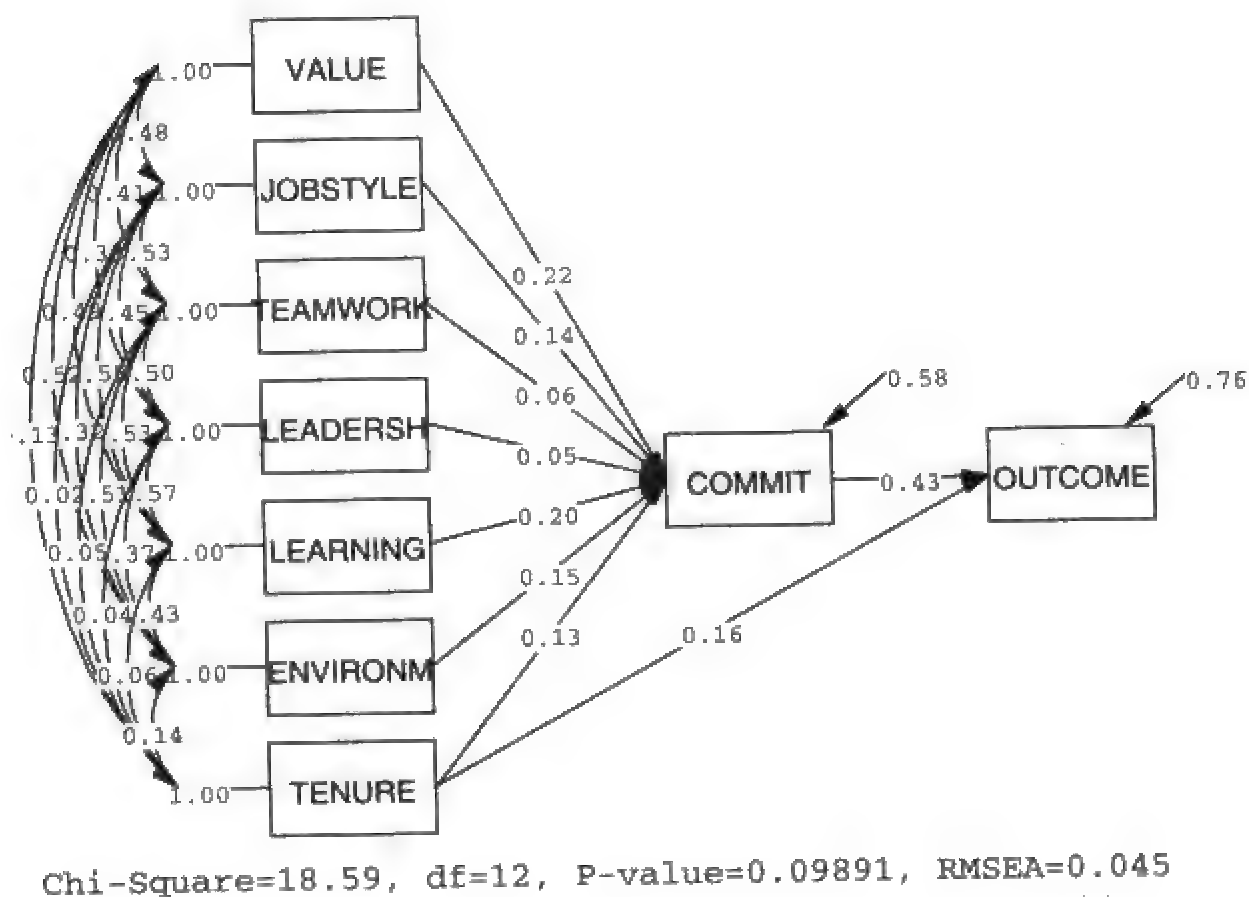


图 7.11 路径分析的标准化终解路径图

在这些回归系数矩阵当中，有些数据已经在先前路径系数的数据中列出（例如七个外源变量对于组织承诺的 Gamma 系数），另外有些参数则是并未被纳入估计的参数。其中，甚至有三个参数经过 MI 指数建议应纳入模型当中加以估计（注 8）。

值得注意的是，此地的回归系数矩阵其实就是路径模式的饱和模型所可能出现的所有回归系数。所谓饱和模式乃是包含了所有可能的路径系数（不包括回溯关系的设定）的路径模型。部分路径分析的文献建议，路径分析的第一步，应从饱和模型的检验开始，对所有的路径参数进行检验，再透过修正程序排除不适合的参数，以完成最恰当的模式。如果使用者欲采取这种途径，实则可以从回归系数矩阵的数据进行分析，也有类似功效。

Regression Matrix Y on X (Standardized)

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
OUTCOME	0.09	0.06	0.02	0.02	0.09	0.07
COMMIT	0.22	0.14	0.06	0.05	0.20	0.15

Regression Matrix Y on X (Standardized)

	TENURE
OUTCOME	0.21
COMMIT	0.13

注 8：被检测具有统计意义但是尚未纳入模型当中的参数。

(四) 直接与间接效应

传统路径分析的重点工作之一，是进行变量之间的效应分析，也就是探讨内生变量被外源变量解释的总体效应、直接效应与间接效应的整理与分析。而 SEM 路径分析的最后一个报表，也就是报告整体与间接效应的数据。

在 SEM 分析程序中，直接效应已经在前面的路径系数以 Beta 与 Gamma 结构参数的形式呈现，而 LISREL 则另外整理出总体效应及间接效应，且包括非标准化与标准化数据。在非标准化效应中，并附上显著性检验（t 检验），使研究者可以直接快速且明确说明路径模型的各效应的统计意义。此一特性，成为 SEM 分析用于路径分析的最大优势之一。这些效应的估计数报表罗列于后。

Total and Indirect Effects

Total Effects of X on Y

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
OUTCOME	0.07 (0.02) 3.31	0.04 (0.02) 2.28	0.02 (0.02) 0.88	0.02 (0.02) 0.90	0.06 (0.02) 2.90	0.05 (0.02) 2.44
COMMIT	0.43 (0.12) 3.63	0.24 (0.10) 2.37	0.11 (0.08) 0.88	0.11 (0.12) 0.90	0.36 (0.12) 3.11	0.28 (0.11) 2.56

注 9：总效应的估计数与显著性检验，显示组织价值对于员工绩效的总效应显著。

Total Effects of X on Y

	TENURE
OUTCOME	0.01 (0.00) 3.72
COMMIT	0.02 (0.01) 2.74

注 10: 年资对于员工绩效的总效应的估计数与显著性检验。达显著水平。

Indirect Effects of X on Y

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
OUTCOME	0.07 (0.02) 3.31	0.04 (0.02) 2.28	0.02 (0.02) 0.88	0.02 (0.02) 0.90	0.06 (0.02) 2.90	0.05 (0.02) 2.44
COMMIT	- -	- -	- -	- -	- -	- -

注 11: 组织价值对于员工绩效的间接效应的估计数与显著性检验。

Indirect Effects of X on Y

	TENURE
OUTCOME	0.00 (0.00) 2.59
COMMIT	- -

Total Effects of Y on Y

	OUTCOME	COMMIT
OUTCOME	- -	0.17 (0.02) 7.99
COMMIT	- -	- -

Standardized Total and Indirect Effects

Standardized Total Effects of X on Y

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
OUTCOME	0.09	0.06	0.02	0.02	0.09	0.07
COMMIT	0.22	0.14	0.06	0.05	0.20	0.15

Standardized Total Effects of X on Y

	TENURE
OUTCOME	0.21
COMMIT	0.13

Standardized Indirect Effects of X on Y

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
OUTCOME	0.09	0.06	0.02	0.02	0.09	0.07
COMMIT	-	-	-	-	-	-

Standardized Indirect Effects

	TENURE
OUTCOME	0.05
COMMIT	-

注 12: 组织价值对员工绩效的间接效应, 标准化数值为.09, 为 r_{21} (.21) 与 β_{12} (.43) 的乘积。

Standardized Total Effects of Y on Y

	OUTCOME	COMMIT
OUTCOME	-	0.43
COMMIT	-	-

注 13: 年资对于员工绩效的间接效应, 标准化数值为.05, 为 r_{27} (.13) 与 β_{12} (.43) 的乘积。总效应.21 为间接效果的.05 与 r_{17} (.16) 的总和。

注 12: 组织价值对员工绩效的间接效应, 标准化数值为.09, 为 γ_{21} (.21) 与 β_{12} (.43) 的乘积。

注 13: 年资对于员工绩效的间接效应, 标准化数值为.05, 为 γ_{27} (.13) 与 β_{12} (.43) 的乘积。总效应.21 为间接效果的.05 与 γ_{17} (.16) 的总和。

以第一个测量变量(组织价值)对于第一个因变量(员工绩效)的效应分析为例, 具有.07 的总效应, 标准化总效应为.09, t 值为 3.31, 显著性检验达到显著水平(注 9)。组织价值对于员工绩效之间并无直接效应, 但是具有一个由组织承诺扮演中介效应的间接效应。因为组织价值对组织承诺的结构系数达到显著 ($\gamma_{21}=.22$), 而组织承诺对员工绩效的结构系数亦达显著 ($\beta_{12}=.43$), 这两个系数的乘积, 就是组织价值对于员工绩效的间接效应, 其显著性检验达到显著水平(注 11), 其标准化数值(也就是 γ_{21} 与 β_{12} 的乘积)为.09(注 12)。

表 7.6 路径分析各项效应分解说明

自 变 量		因变量（内生变量）			
		Y ₂ 组织承诺		Y ₁ 员工绩效	
		标准化效应	t 值	标准化效应	t 值
外 源 变 量	X ₁ 组织价值				
	直接效应	.22	3.63***	-	-
	间接效应	-	-	.09	3.31***
	总效应	.22	3.63***	.09	3.31***
	X ₂ 工作方式				
	直接效应	.14	2.37**	-	-
	间接效应	-	-	.06	2.28*
	总效应	.14	2.37**	.06	2.28*
	X ₃ 团队合作				
	直接效应	.06	0.88	-	-
	间接效应	-	-	.02	0.88
	整体效应	.06	0.88	.02	0.88
	X ₄ 领导风格				
	直接效应	.05	0.90	-	-
	间接效应	-	-	.09	0.90
	总效应	.05	0.90	.09	0.90
	X ₅ 学习成长				
	直接效应	.20	3.11**	-	-
	间接效应	-	-	.07	2.90**
	总效应	.20	3.11**	.07	2.90**
	X ₆ 环境气氛				
	直接效应	.15	2.56*	-	-
	间接效应	-	-	.05	2.44**
	总效应	.15	2.56*	.05	2.44**
	X ₇ 年资				
	直接效应	.13	2.74**	.16	2.89**
	间接效应	-	-	.05	2.59**
	总效应	.13	2.74**	.21	3.72***
内 生 变 量	Y ₂ 组织承诺				
	直接效应			.43	7.99***
	间接效应			-	-
	总效应			.43	7.99***

注：t 值大于 1.96 时， $p < .05$ ；大于 2.58 时， $p < .01$ ；大于 3.29 时， $p < .001$

总效应为直接效应与间接效应的总和，因此第一个测量变量（组织价值）对于第一个因变量（员工绩效）的效应分析如下：

$$\text{总效应} = \text{直接效应} + \text{间接效应}$$

$$0.09 = 0.00 + 0.09$$

再以年资为例，年资作为控制变量，对于员工绩效与组织承诺均具有显著的直接效应，分别为 r_{17} (.16) 与 r_{27} (.13)。与此同时，组织承诺对于员工绩效又有显著的直接效应 ($\beta_{12}=.43$)，因此，年资变量对于员工绩效即具有一个.05 的间接效应，其数值为 r_{27} (.13) \times β_{12} (.43) 之乘积（注 13）。年资对于员工绩效的效应分析如下：

$$\text{总效应} = \text{直接效应} + \text{间接效应}$$

$$0.21 = .16 + 0.05$$

SEM 的路径分析不但可以依上述公式得到总效应的数值，更可以进行显著性检验，给予研究者相当大的便利，使得最后的路径分析整理表（表 7.5），不但容易编制，更具有显著性检验结果来证明其统计意义。这就是为何有越来越多的研究者，舍弃传统的回归分析，改以 SEM 取向进行路径分析。

六、模型修饰

前述参数估计与残差分析结果指出，有两个被估计的结构参数未达显著，而有个数未被估计的参数残差值偏高，显示这些模型有进一步修正的空间。以下，我们透过模型修饰指数的检阅，找出可以进行修正的参数。

其中最值得注意的是员工绩效对组织承诺的结构参数 β_{21} ， MI 值为 14.92（注 14），表示此参数应纳入模型当中，然而若该参数纳入估计，表示路径模型具有回馈循环效应，也就是员工绩效与组织承诺两个外源变量互为直接效应，即员工绩效与组织承诺之间具有回馈循环效应。

Modification Indices and Expected Change

Modification Indices for BETA			
	OUTCOME	COMMIT	
OUTCOME	---	---	
COMMIT	14.92	---	注 14: MI 值大于 5, 表示此参数宜纳入模型中。此参数若纳入模型, 表示具有回馈循环效应。
Expected Change for BETA			
	OUTCOME	COMMIT	
OUTCOME	---	---	
COMMIT	-0.80	---	
Standardized Expected Change for BETA			
	OUTCOME	COMMIT	
OUTCOME	---	---	
COMMIT	-0.80	---	注 15: 预期改变量为负值, 表示两个变量变动方向相反。

由预期改变量来看, 结构参数 β_{21} 的效应将为负值 (注 15), 表示员工绩效越高, 组织承诺将会越低。此一结果从理论逻辑性来看, 实难解释。因此, 即使 MI 指数建议该参数可以纳入模型当中, 但是基于理论的合理性, 纳入反而造成解释上的困难。

进一步检验 Gamma 参数的 MI 指数, 发现有数个参数的回归系数皆有统计的意义。例如学习成长、领导风格、组织价值等变量对员工绩效的预测, MI 值分别为 12.73、8.89、6.86, 显示这些变量是具有解释力的变量。同时预期改变量均为正值, 显示这些变量对于员工绩效的预测值将为正值。这些变量得分越高, 员工绩效表现越佳, 具有理论上的合理性。

Modification Indices for GAMMA

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
OUTCOME	6.86	3.34	6.53	8.89	12.73	3.56
COMMIT	-	-	-	-	-	-

Expected Change for GAMMA

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
OUTCOME	0.13	0.07	0.11	0.14	0.16	0.08
COMMIT	-	-	-	-	-	-

Standardized Expected Change for GAMMA

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
OUTCOME	0.16	0.11	0.15	0.17	0.22	0.11
COMMIT	-	-	-	-	-	-

有关残差部分的修饰指数则指出, 员工绩效与组织承诺的残差间有显著的修正空间, MI 值为 14.92, 此一数值与先前之间 β_{21} 的 MI 值都很高。也就是说, 在员工绩效与组织承诺之间, 存有一定的可估计共变, 处理方式可以 Beta 参数来估计, 或以残差项的相关来处理均可。然而, 前面对于 β_{21} 预期该变量为负值, 较具理论与实务的合理性。

从预期改变量来看, 增加员工绩效与组织承诺的残差的相关系数估计, 也将带来负值 (-0.25), 反映的是这两个内生变量的被解释剩余共变异量为负值。

Modification Indices for THETA-EPS

	OUTCOME	COMMIT
OUTCOME	-	-
COMMIT	14.92	-

Expected Change for PSI

	OUTCOME	COMMIT
OUTCOME	-	-
COMMIT	-0.25	-

Standardized Expected Change for PSI

	OUTCOME	COMMIT
OUTCOME	-	-
COMMIT	-0.24	-

其他残差修饰指数显示，年资与组织承诺的残差之间具有修正的价值，MI 为 11.78（注 16）。但是，尽管修正指数如此高，由于路径分析的一项基本假设是解释残差与外源变量之间应为相互独立，因此若要纳入此一参数将违反独立性假设。

图 7.12 整理出所有基于 MI 指数所建议增加的参数，并以路径图的方式呈现，有助于模型修饰决定。

基于上述的考虑，模型修饰得以增加一条由“学习成长”到“员工绩效”的直接效应。亦即增加 γ_{15} 。LISREL 语法中新增 FREE GA (1,5)，预期回归系数数值为.16，能降低的卡方值约为 12.73。

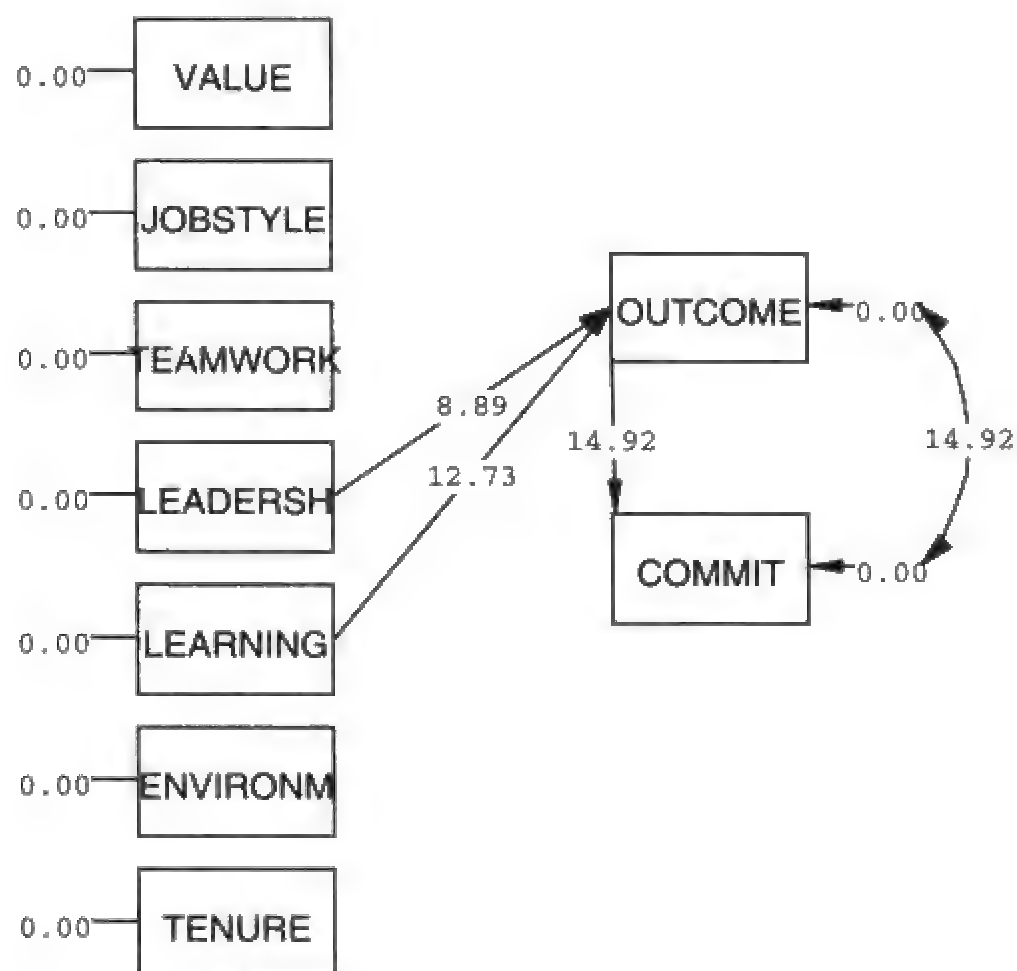
Modification Indices for THETA-DELTA-EPS			
	OUTCOME	COMMIT	
	-----	-----	
VALUE	2.00	2.00	
JOBSTYLE	0.13	0.13	
TEAMWORK	0.48	0.48	
LEADERSH	1.58	1.58	
LEARNING	3.27	3.27	
ENVIRONM	0.01	0.01	
TENURE	0.75	11.78	注 16: 年资与组织承诺残差具有修正价值，但是此一相关系数的估计将违反独立性假设。

Expected Change for THETA-DELTA-EPS		
	OUTCOME	COMMIT
	-----	-----
VALUE	0.03	-0.17
JOBSTYLE	-0.01	0.05
TEAMWORK	0.01	-0.08
LEADERSH	0.03	-0.15
LEARNING	0.04	-0.22
ENVIRONM	0.00	-0.01
TENURE	2.11	20.12

■ 修饰后的结果

如预期般的，经过增加一条由“学习成长”到“员工绩效”的直接效应后，整体拟合大幅提升，卡方值降至 5.81 ($df=11, p=.886$)， $RMSEA=0.00$ ，近似于完美拟合。“学习成长”到“员工绩效”的直接效应为.22，比原先预期的.16 还来得高，如图 7.13。两

个模型为嵌套模型，因此可以进行卡方差异检验， $\Delta\chi^2=12.78$ ， $\Delta df=1$ ，大于自由度为 1 的 χ^2 临界值 3.89，显示此一修饰具有统计的意义。因此我们的结论为：“组织的学习成长气氛越浓，员工的绩效表现越好 (.22)”。从间接效应来看，“学习成长气氛会透过影响员工的组织承诺 (.20)，进而影响绩效 (.31)”，间接效应为.06，直接效应为.22，合计总效应为.28。



Chi-Square=18.59, df=12, P-value=0.09891, RMSEA=0.045

图 7.12 外显变量路径分析的修正建议路径图

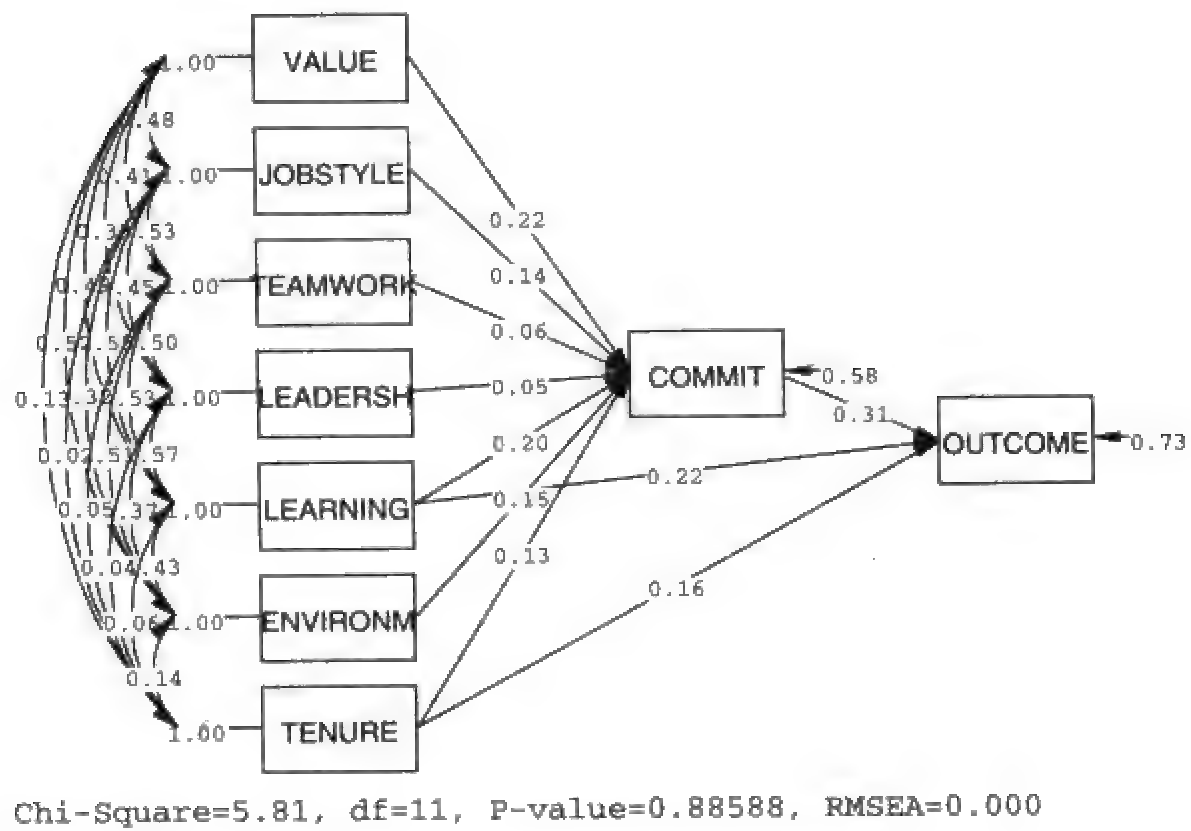


图 7.13 路径分析修正后终解图

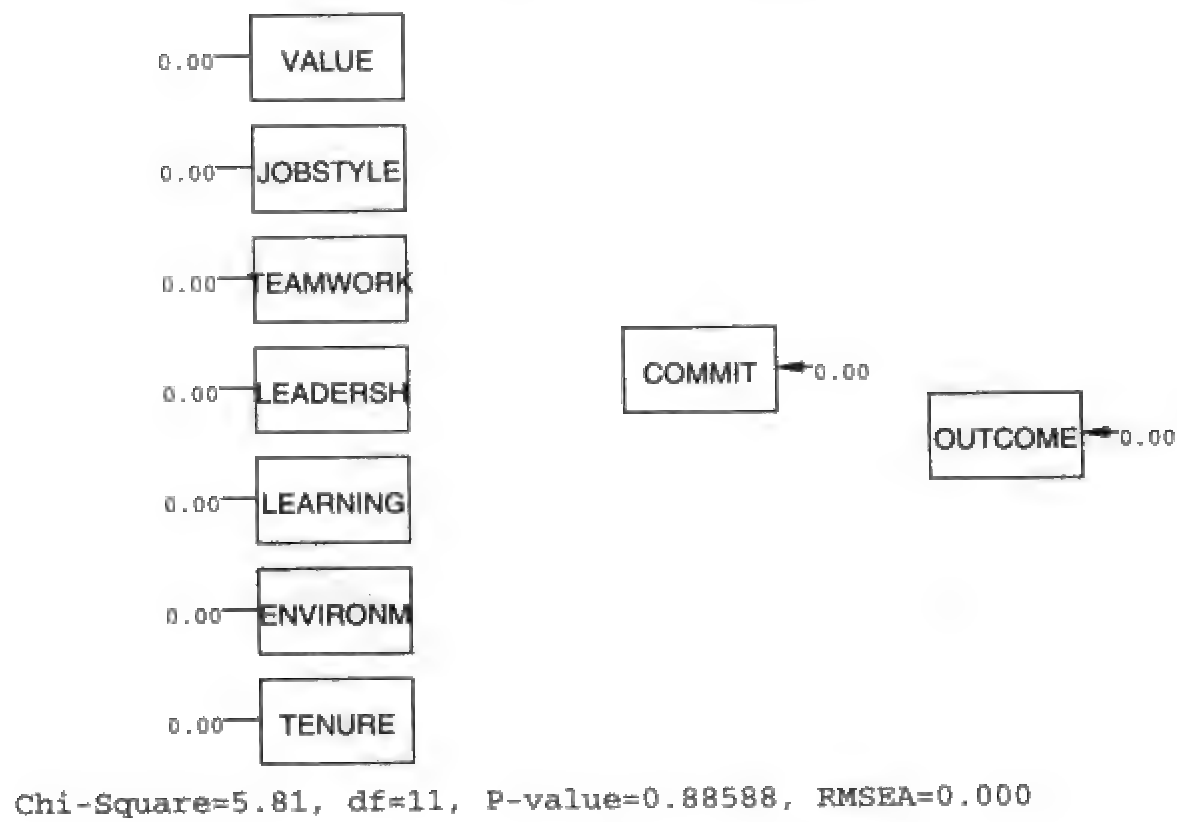


图 7.14 修正后修正指数数据图

有趣的是,原模型的 MI 指数指出四个需要修饰的参数,但是经过新增一条参数后,新模型的 MI 则无一达到需要修饰的水平(见图 7.14)。这说明了修饰指数所建议的单变量参数;然而一旦某个参数改变后,修饰指数内容即也随之改变。另一方面,这反映了模型修饰的困难与吊诡,因为各参数的调整是“牵一发而动全身”。除了仰赖统计测量的证据,也依靠理论与文献支持何者最适合于修饰,更重要的,或许还需要研究者的直觉与主观判断。那么,SEM 究竟是客观科学还是一种辩证艺术呢?这确实值得人们深思。

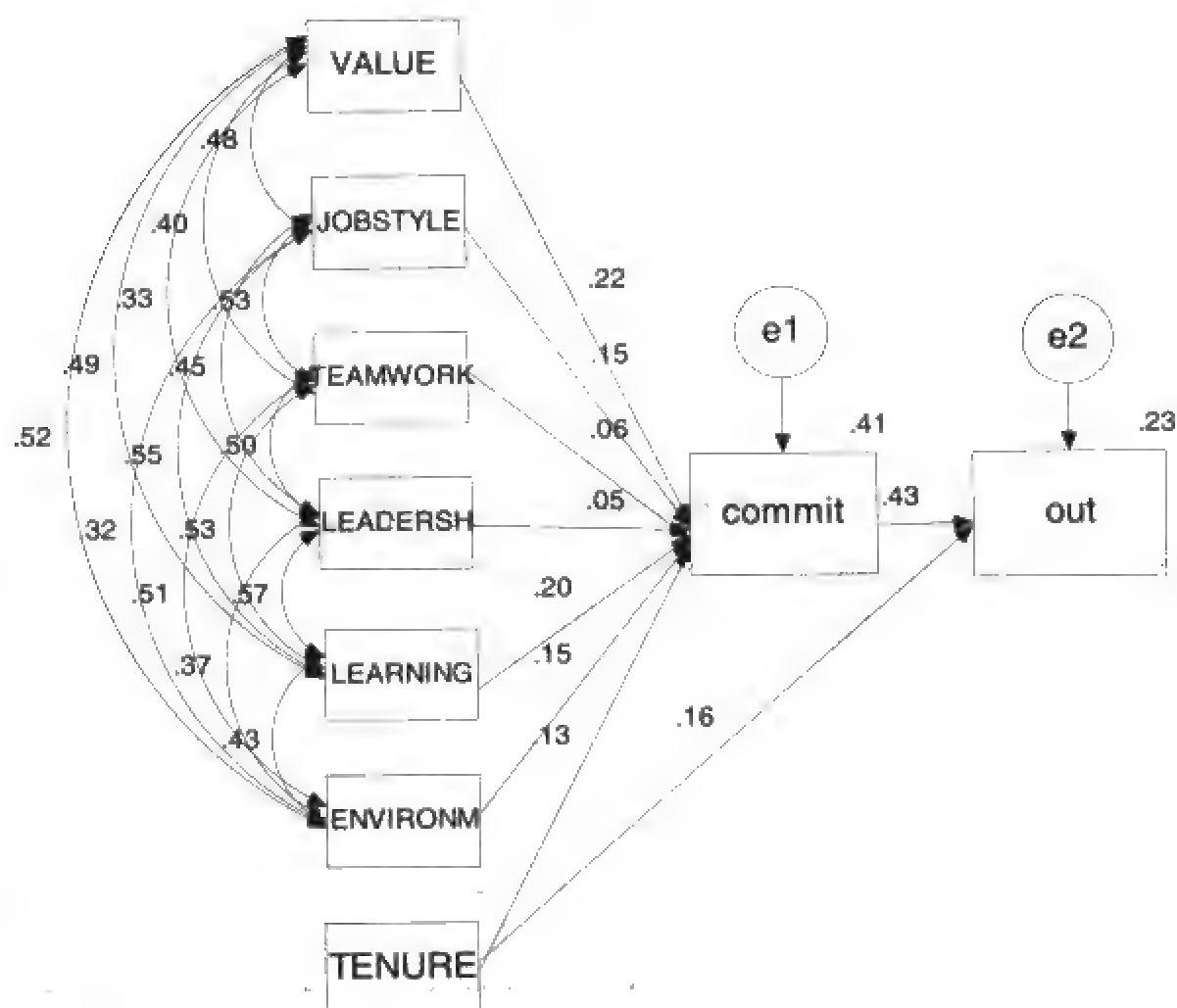
表 7.7 路径模型修饰前后的模型拟合度比较表

Model	卡方检验量		df	RMSEA	NNFI	CFI	GFI	SRMR
	χ^2	WLS χ^2						
修饰前	19.24 $p=.083$	18.59 $p=.099$	12	.045	.99	1.00	.99	.045
修饰后	5.87 $p=.88$	5.81 $p=.89$	11	.000	1.01	1.00	1.00	.018
差异	13.37 $p<.05$	12.78 $p<.05$	1					

第四节 AMOS 的路径分析

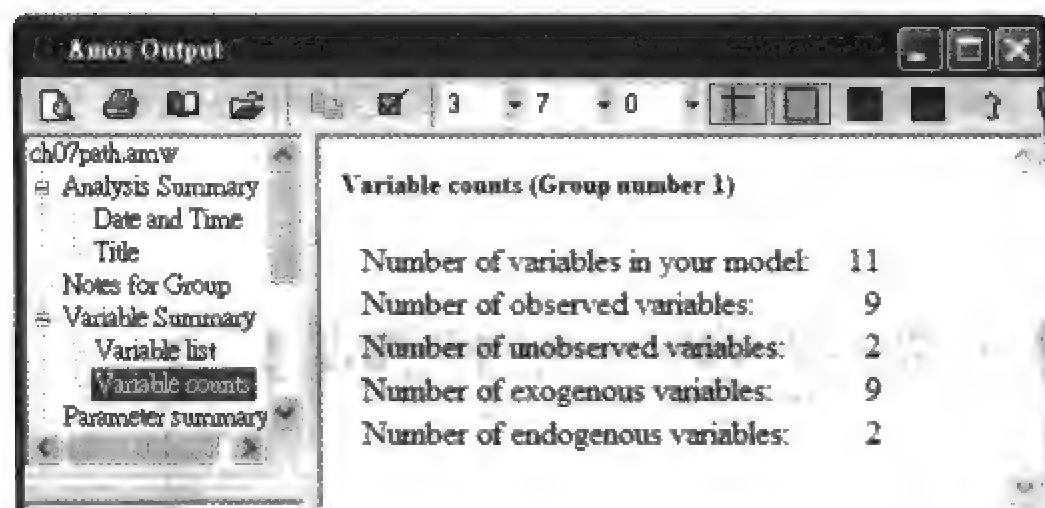
一、模型界定

依照本范例的假设模型,绘制图形,下图为执行完毕之后的标准化最终解路径图。

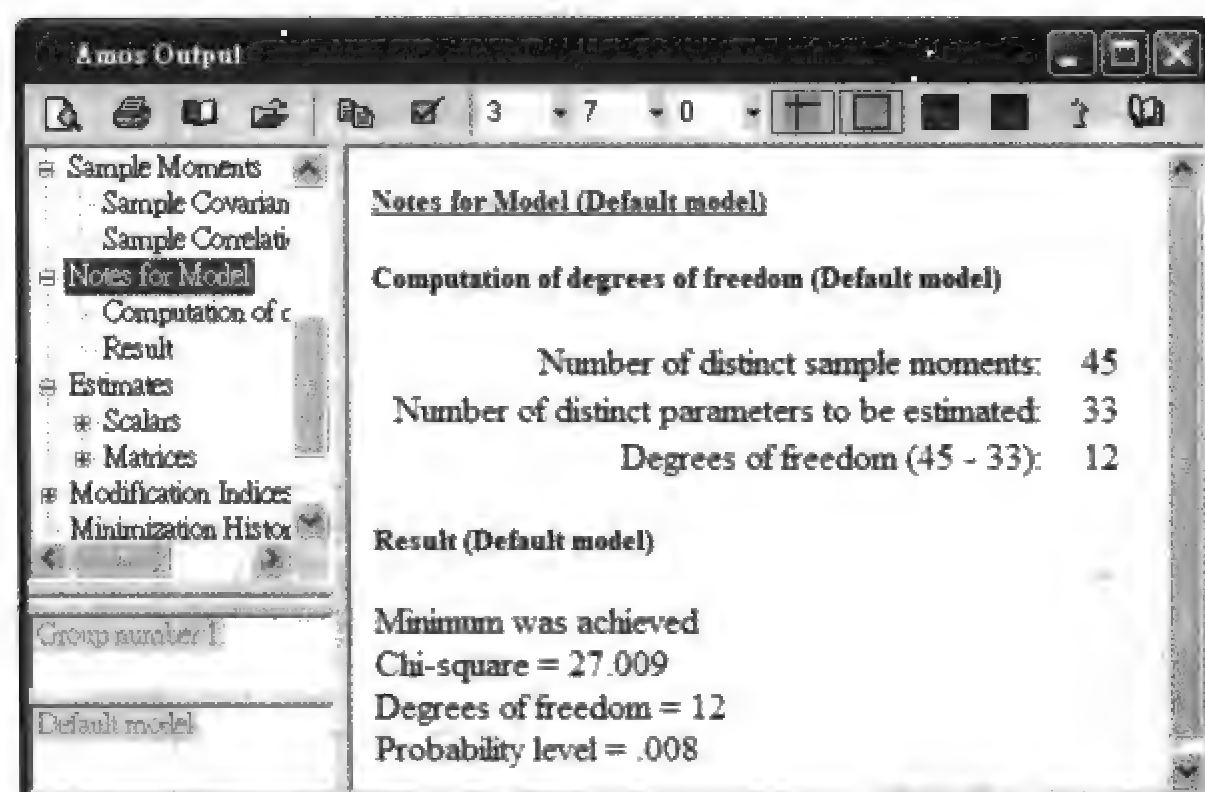


二、AMOS 的报表解读

本范例已事先将遗漏值进行处理，样本数共计 281 个。变量的摘要显示模型中的所有变量，包含观察变量（内生变量）与非观察变量（外源变量），观察变量有 9 个，非观察变量有 2 个，共计 11 个变量。

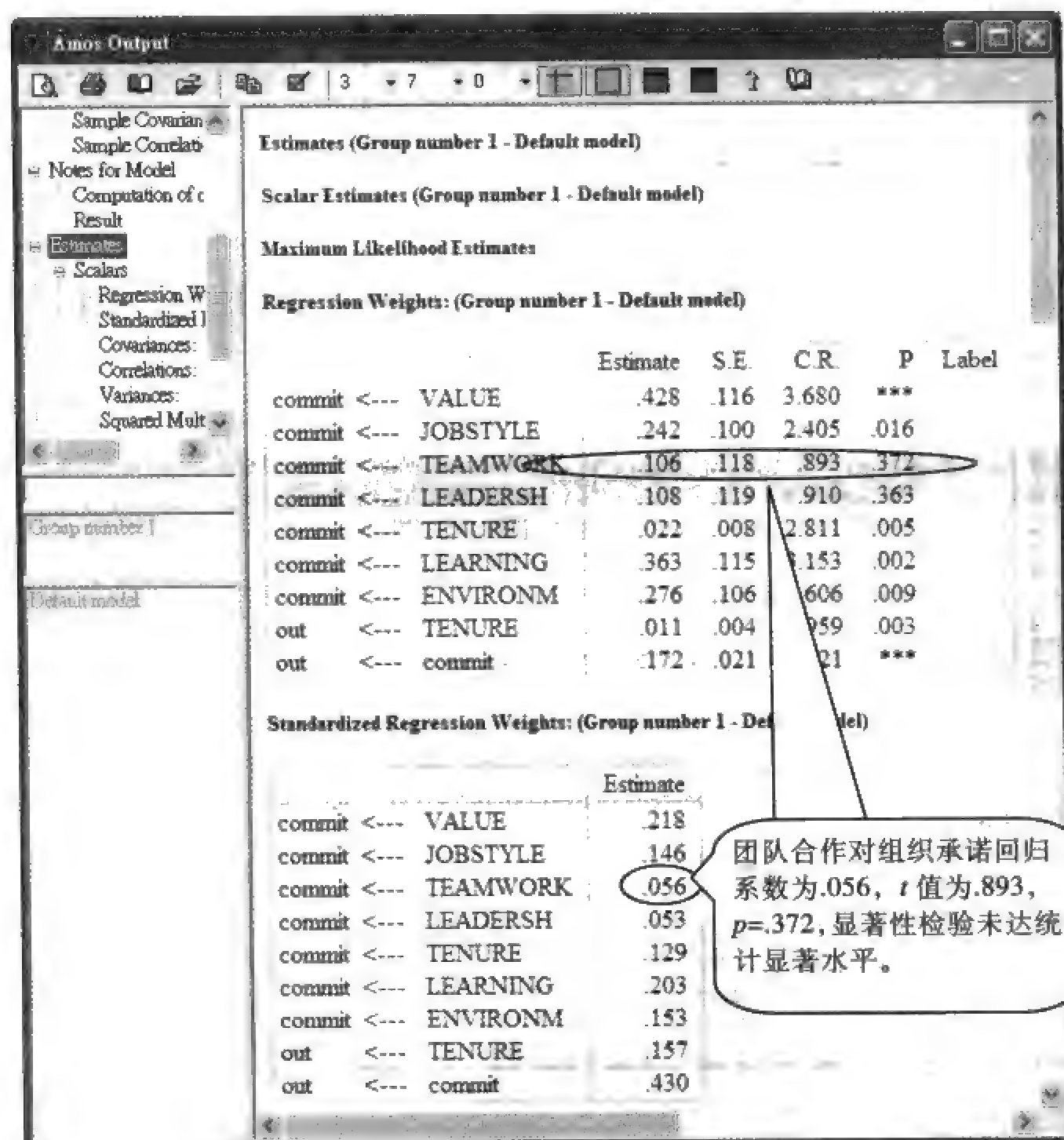


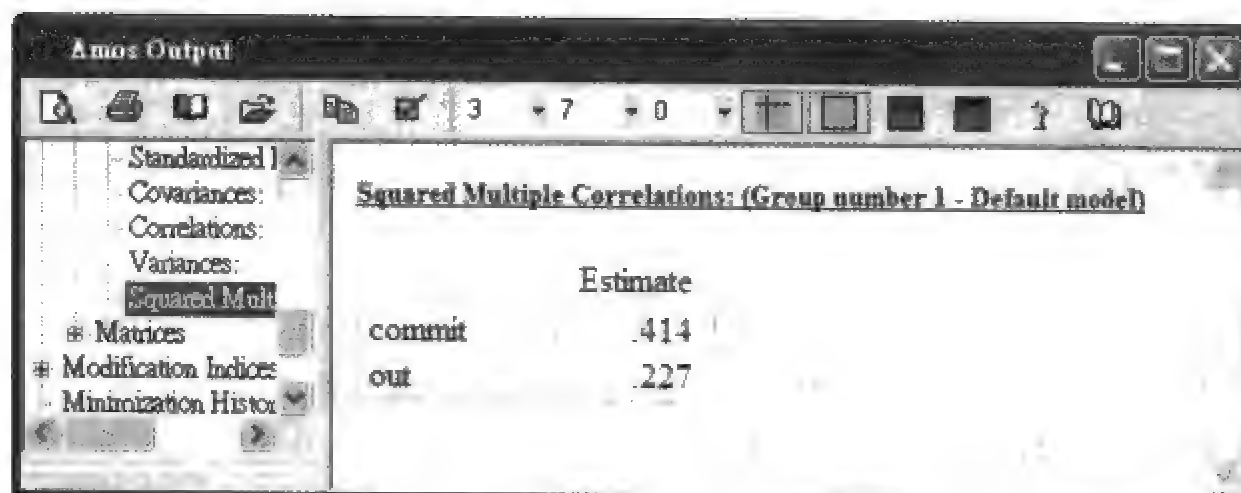
模型的纪录显示模型的测量数据数为 45 个, 有 33 个参数被估计, 因此自由度为 12。卡方值=27.009, $p=.008$ 。



从参数估计结果的报表数据可以得知, TEAMWORK (团队合作) 与 LEADERSH (领导风格) 的结构参数未达显著, 标准化回归系数值分别为.056 与.053, 表示这

两条参数对于 commit（组织承诺）的解释不具有统计的意义。从 Covariances 的数据显示, 各外源变量的方差与协方差皆达显著水平。从 Squared Multiple Correlations (SMC) 的估计结果, 显示两个内生变量（组织承诺与员工绩效）被外源变量解释的比例, 分别为.414 与.227。

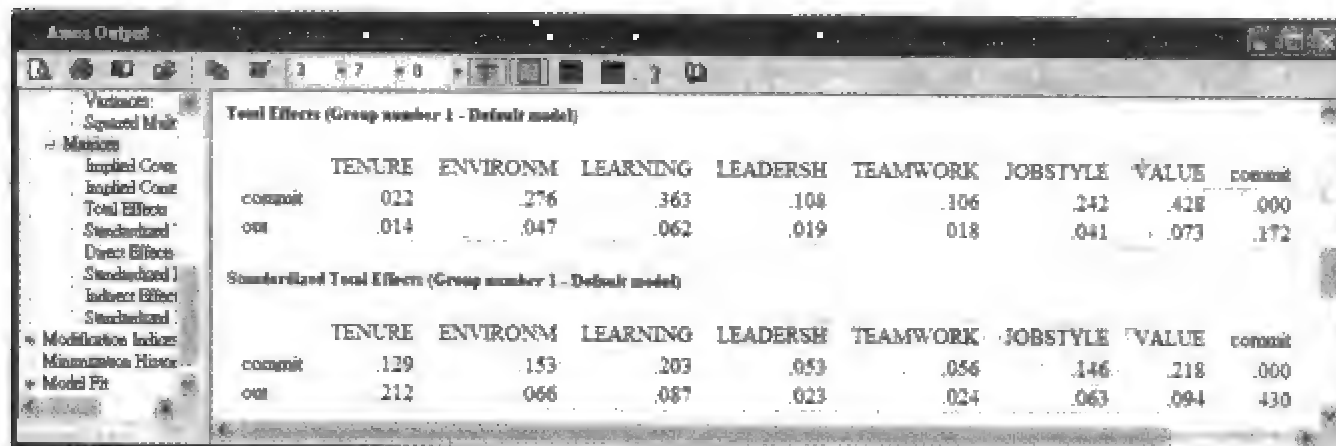




Squared Multiple Correlations: (Group number 1 - Default model)

	Estimate
commit	.414
out	.227

从变量之间的效应分析数据来看, 我们应报告标准化的总体效应、标准化的直接效应以及标准化的间接效应。标准化的直接效应已在前面的结构参数中呈现, AMOS 报表另外整理出总体效应与间接效应。唯一可惜的是, 报表中并未附上显著性检验 (t 检验), 使用者必须另外利用公式自行计算。以下仅列出部分的报表。完整报表请参阅档案 ch07.Amos Output。

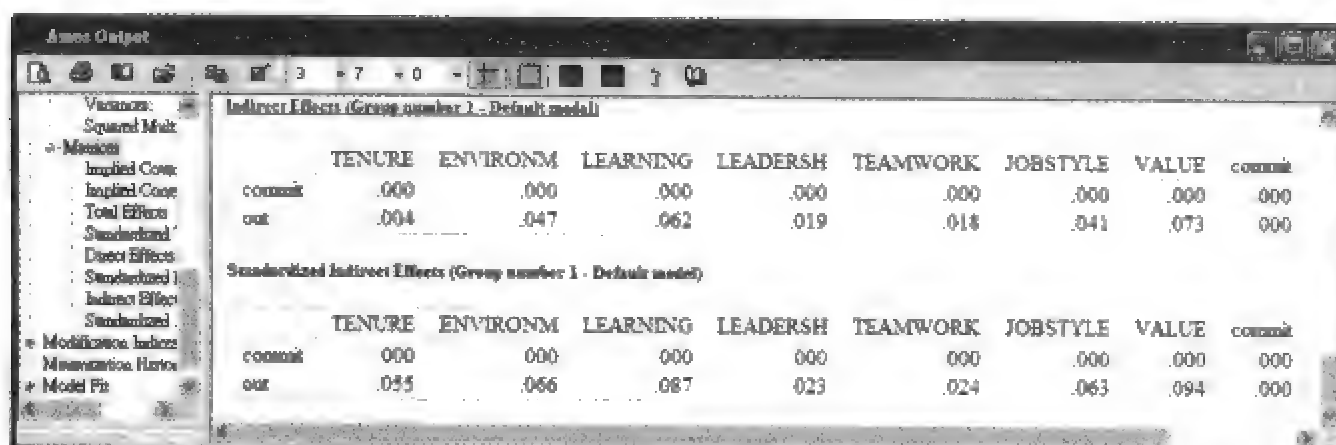


Total Effects (Group number 1 - Default model)

	TENURE	ENVIRONM	LEARNING	LEADERSH	TEAMWORK	JOBSTYLE	VALUE	commit
commit	.022	.276	.363	.108	.106	.242	.428	.000
out	.014	.047	.062	.019	.018	.041	.073	.172

Standardized Total Effects (Group number 1 - Default model)

	TENURE	ENVIRONM	LEARNING	LEADERSH	TEAMWORK	JOBSTYLE	VALUE	commit
commit	.129	.153	.203	.053	.056	.146	.218	.000
out	.212	.066	.087	.023	.024	.063	.094	.430



Indirect Effects (Group number 1 - Default model)

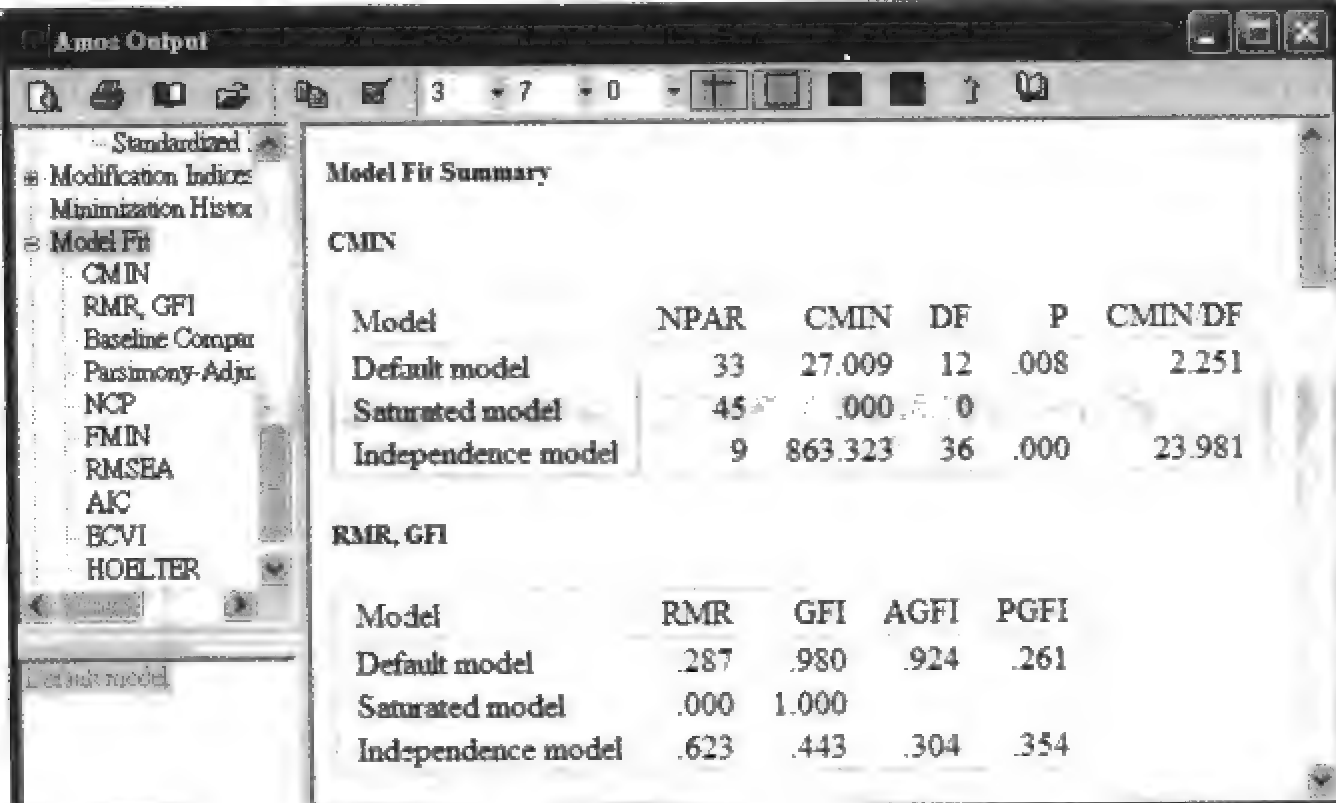
	TENURE	ENVIRONM	LEARNING	LEADERSH	TEAMWORK	JOBSTYLE	VALUE	commit
commit	.000	.000	.000	.000	.000	.000	.000	.000
out	.004	.047	.062	.019	.018	.041	.073	.000

Standardized Indirect Effects (Group number 1 - Default model)

	TENURE	ENVIRONM	LEARNING	LEADERSH	TEAMWORK	JOBSTYLE	VALUE	commit
commit	.000	.000	.000	.000	.000	.000	.000	.000
out	.055	.066	.087	.023	.024	.063	.094	.000

三、模型拟合度分析

从模型拟合度摘要表中显示 $CMIN/DF$ 为 2.251, $p=.008$, $SRMR=.060$, $RMSEA=.067$, TLI ($NNFI$)、 NFI 、 CFI 、 GFI 皆大于 .90, 整体而言, 本模型拟合度堪称理想。因此可以直接报告最后的研究数据。



The screenshot shows the 'Amos Output' window with the 'Model Fit Summary' section expanded. The left sidebar lists various fit indices, and the main area displays two tables: 'CMIN' and 'RMR, GFI'.

CMIN					
Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	33	27.009	12	.008	2.251
Saturated model	45	.000	0		
Independence model	9	863.323	36	.000	23.981

RMR, GFI				
Model	RMR	GFI	AGFI	PGFI
Default model	.287	.980	.924	.261
Saturated model	.000	1.000		
Independence model	.623	.443	.304	.354

第五节 Mplus 的路径分析

一、Mplus 语法

```
TITLE: Ch7a Path analysis
DATA: FILE IS ch7.dat;
VARIABLE: NAMES ARE OUTCOME COMMIT VALUE JOBSTYLE TEAMWORK LEADERSH LEARNING
          ENVIRONM TENURE;
MODEL:
  COMMIT on VALUE JOBSTYLE TEAMWORK LEADERSH LEARNING ENVIRONM TENURE;
  OUTCOME on COMMIT TENURE;
MODEL INDIRECT:
  OUTCOME IND COMMIT VALUE;
  OUTCOME IND COMMIT JOBSTYLE;
  OUTCOME IND COMMIT TEAMWORK;
  OUTCOME IND COMMIT LEADERSH;
  OUTCOME IND COMMIT LEARNING;
  OUTCOME IND COMMIT ENVIRONM;
  OUTCOME IND COMMIT TENURE;
  OUTCOME IND TENURE;
OUTPUT:
  STANDARDIZED;;
```

间接效应估计
主指令为 Model Indirect:
要求计算机列出间接效应及其标准误。
最终因变量放在 IND 的左边，最初自变量放在同一行的最右边。外源变量仅能放在最右侧。

注意事项
为了计算整体效应，除了必须指明间接作用之外，还需指定直接效应。各外源变量中，仅有 TENURE 对 OUTCOME 有直接与间接作用。

二、Mplus 报表

TESTS OF MODEL FIT		
Chi-Square Test of Model Fit		
Value		19.320
Degrees of Freedom		6
P-Value		0.0037
Chi-Square Test of Model Fit for the Baseline Model		
Value		250.276
Degrees of Freedom		15
P-Value		0.0000
CFI/TLI		
CFI		0.943
TLI		0.858

Loglikelihood

HO Value	-3518.560
HI Value	-3508.900

Information Criteria

Number of Free Parameters	13
Akaike (AIC)	7063.120
Bayesian (BIC)	7110.419
Sample-Size Adjusted BIC ($n^* = (n + 2) / 24$)	7069.196

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.089	
90 Percent C.I.	0.046	0.135
Probability RMSEA $\leq .05$	0.064	

SRMR (Standardized Root Mean Square Residual)

Value	0.041
-------	-------

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
COMMIT ON				
VALUE	0.428	0.116	3.679	0.000
JOBSTYLE	0.243	0.100	2.418	0.016
TEAMWORK	0.105	0.118	0.891	0.373
LEADERSH	0.108	0.119	0.911	0.362
LEARNING	0.363	0.115	3.154	0.002
ENVIRONM	0.276	0.106	2.598	0.009
TENURE	0.022	0.008	2.780	0.005
OUTCOME ON				
COMMIT	0.172	0.021	8.118	0.000
TENURE	0.011	0.004	2.925	0.003
Intercepts				
OUTCOME	2.488	0.200	12.453	0.000
COMMIT	2.537	0.539	4.703	0.000
Residual Variances				
OUTCOME	0.306	0.026	11.853	0.000
COMMIT	1.456	0.123	11.853	0.000

STANDARDIZED MODEL RESULTS

STDYX Standardization

直接效应

原始估计数与标准化估计数 (STDXY)。
一般论文所呈现的均为 STDXY, 即为标
准化回归系数。

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
COMMIT ON				
VALUE	0.217	0.058	3.716	0.000
JOBSTYLE	0.145	0.060	2.427	0.015
TEAMWORK	0.055	0.062	0.891	0.373
LEADERSH	0.053	0.058	0.912	0.362
LEARNING	0.201	0.063	3.178	0.001
ENVIRONM	0.152	0.058	2.609	0.009
TENURE	0.128	0.046	2.788	0.005
OUTCOME ON				
COMMIT	0.432	0.048	8.931	0.000
TENURE	0.155	0.053	2.952	0.003
Intercepts				
OUTCOME	3.931	0.431	9.128	0.000
COMMIT	1.596	0.385	4.144	0.000
Residual Variances				
OUTCOME	0.763	0.044	17.227	0.000
COMMIT	0.576	0.045	12.871	0.000

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
OUTCOME	0.237	0.044	5.343	0.000
COMMIT	0.424	0.045	9.477	0.000

 R^2

两个因变量被直接效应解释的百分比。

TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Effects from TENURE to OUTCOME				
Total	0.014	0.004	3.768	0.000
Total indirect	0.004	0.001	2.630	0.009
Specific indirect				
OUTCOME				
COMMIT				
TENURE	0.004	0.001	2.630	0.009
Direct				
OUTCOME				
TENURE	0.011	0.004	2.925	0.003

……(略)

STANDARDIZED TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

STDYX Standardization

标准化间接效应与总效应

此为标准化估计数，等于标准化回归系数。

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Effects from TENURE to OUTCOME				
Total	0.211	0.055	3.847	0.000
Total indirect	0.055	0.021	2.677	0.007
Specific indirect				
OUTCOME COMMIT TENURE	0.055	0.021	2.677	0.007
Direct				
OUTCOME TENURE	0.155	0.053	2.952	0.003
Effects from VALUE to OUTCOME				
Sum of indirect	0.093	0.027	3.406	0.001
Specific indirect				
OUTCOME COMMIT VALUE	0.093	0.027	3.406	0.001
Effects from JOBSTYLE to OUTCOME				
Sum of indirect	0.063	0.027	2.334	0.020
Specific indirect				
OUTCOME COMMIT JOBSTYLE	0.063	0.027	2.334	0.020
Effects from TEAMWORK to OUTCOME				
Sum of indirect	0.024	0.027	0.887	0.375
Specific indirect				
OUTCOME COMMIT TEAMWORK	0.024	0.027	0.887	0.375

Effects from LEADERSH to OUTCOME				
Sum of indirect	0.023	0.025	0.906	0.365
Specific indirect				
OUTCOME COMMIT LEADERSH	0.023	0.025	0.906	0.365
Effects from LEARNING to OUTCOME				
Sum of indirect	0.087	0.029	2.976	0.003
Specific indirect				
OUTCOME COMMIT LEARNING	0.087	0.029	2.976	0.003
Effects from ENVIRONM to OUTCOME				
Sum of indirect	0.065	0.026	2.496	0.013
Specific indirect				
OUTCOME COMMIT ENVIRONM	0.065	0.026	2.496	0.013
Effects from TENURE to OUTCOME				
Sum of indirect	0.055	0.021	2.677	0.007
Specific indirect				
OUTCOME COMMIT TENURE	0.055	0.021	2.677	0.007

第六节 结 语

路径分析可以说是当代社会科学研究当中最重要的统计技术之一。从 20 世纪 70 年代以来，路径分析的相关论文如同雨后春笋般出现，所讨论的议题涉及了方法学、分析技术、软件应用等不一而足，所牵涉的领域则扩及社会与行为科学的所有领域。Wolfe (2003) 即撰文指出，这些有关路径分析的珍贵讨论，从发展之初到当代，散布在各个领域之间，实在难以统整，亦无法在一篇论文或一本专著中加以

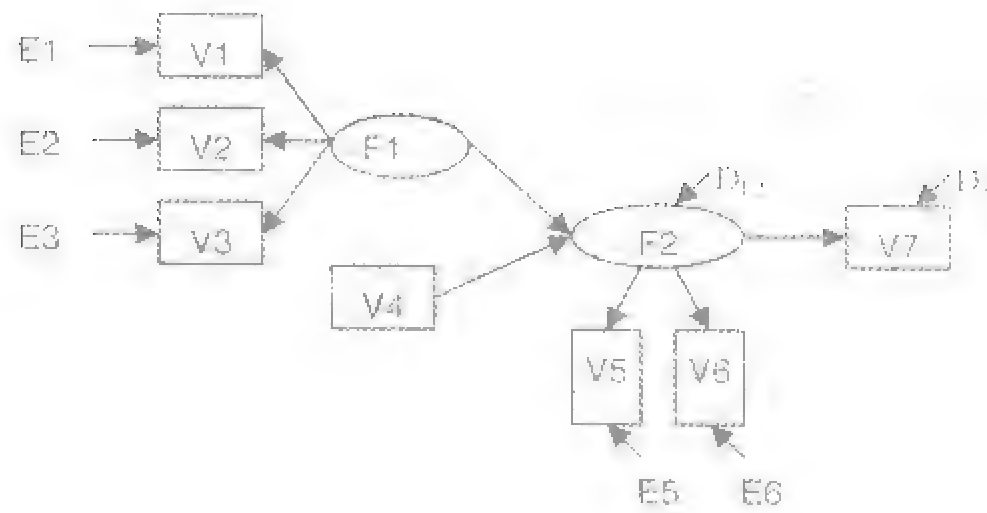
涵盖。因此，Wolfe (2003) 在《结构方程模型》第 10 卷本当中发表了一篇路径分析相关议题的大事记，罗列了有关路径分析各个主题的重要文献出处，也算是替路径分析作了一个大整合。但是，就如同 Wolfe 所说的，路径分析仍在持续成长中，最近的议题包括了多层次的测量、纵贯性的成长测量、类别变量与非正态化数据的分析、估计标准误的不偏估计、遗漏值处理问题，而这些论文亦散见于各领域的相關期刊当中。显见路径分析不但是深受各方重视的分析技术，学者对其发展的掌握也格外的辛苦。

尽管 SEM 分析对于路径分析提供了相当便利的分析工具，但是传统取向的路径分析仍然相当普遍。一篇最近的回顾文献则指出，在当代 (1993-1997) 以 SEM 取向进行的路径分析研究当中，仍有将近四分之一的论文是采取没有潜在变量的路径分析设计，也就是路径分析的分析策略 (Maccallum & Austin, 2000)。作者认为显然仍有相当多的研究人员不明瞭潜在变量在路径分析当中的重要性，同时这些研究背后可能潜藏着许多测量与统计上的误差问题。显然 Maccallum 与 Austin 偏好潜在变量路径分析的应用，但是，事实上在许多研究领域，例如消费研究、社会调查，研究者所处理的数据多为具体、可观察的真实测量，研究中并没有潜在变量的概念，因此自然没有必要使用到潜在变量路径分析的分析策略。也即是基于此一观点，本书同时介绍了路径分析与潜在变量路径分析两种分析范式，使用者可以视需要选择其一来进行相关的数据分析，但是也应注意 Maccallum 与 Austin (2000) 所提醒的路径分析策略可能存在的测量误差问题。

不论路径分析或潜在变量路径分析这两种分析策略何者为优，一个显著的事实是以多次多元回归来进行路径分析的传统策略已经面临被淘汰的命运，改用 SEM 分析软件进行路径分析已经是不可避免的趋势。除非研究者的研究数据不涉及路径关系的探讨，仍可以利用多元回归分析处理一组自变量与一个因变量之间的关系讨论。其他有关中介变量、调节变量乃至多重因变量的回归分析，都将以 SEM 模式为标准程序。这也说明了 SEM 技术在未来量化研究范式的重要性将会不断提高。

第八章

结构方程模型：统合模型分析



同时带有测量模型与结构模型的结构方程模型分析称为统合模型分析，可以说是结构方程模型各种基本模型的整合运用，也可以视为路径分析与验证性因素分析的综合分析，又称为结构回归模型（structural regression models）（Raykov & Marcoulides, 2000）。

当验证性因素分析与路径分析的概念在结构方程模型分析过程中相融合时，因素分析的潜在构念分析技术恰好将路径分析所忽略的抽象构念的测量与误差的处理问题予以补足，进而强化了路径模型的内在测量基础。另一方面，路径分析所擅长的因果关系结构分析，弥补了因素分析只能用来探讨测量变量与潜在构念内在关系的遗憾，将因素分析的触角伸向了潜在构念的外部结构关系，因此是 SEM 分析当中最重要的一般化模式。

第一节 统合模型的基本概念

一、路径分析与因素分析模型的整合

自从 SEM 发展以来，路径分析可以轻易地在 SEM 中加以模块与检验，因此路径分析逐渐改以 SEM 的方式来处理。SEM 取向的路径分析的主要特色是可以利用测量变量间的共变情形，同时估计模型参数，并可配合研究者所提出的特定假设模型检验理论模型与观察数据的適切性，找出最佳的模型。此外，SEM 取向路径分析的另一个优势是可以处理潜在变量的估计，换言之，SEM 取向的路径分析可以在处理潜在变量的同时，也进行路径因果关系的检测。也正因如此，SEM 取向的路径分析还可以处理测量误差的控制。

在传统的路径分析中，用来解释或预测其他变量的解释性变量，通常被假设是没有测量误差，或其测量误差可以被忽略。仅有被解释或被预测的变量被解释残差可以被估计出来。在 SEM 取向的路径分析中，不论是潜在变量或观察变量，测量误

差都可以被有效地估计并排除在分析过程之外。因此，SEM 取向的路径分析有取代传统回归取向路径分析的诸多优势。

从路径分析的原理来看，统合模型的结构化假设可以假定各种直接或间接因果关系，并利用模型拟合度分析技术进行参数估计，找出路径系数并检验其显著性；从因素分析的角度来看，统合模型的测量模型则可以用来检测一群测量变量背后的潜在因素，验证研究者提出的假设模型，并以竞争模型的比较寻找最佳的因素结构。

在测量的内在基础与外部的结构关系皆能兼顾的优势下，统合模型分析一举整合了路径分析与因素分析的核心概念，回避了这些传统多变量统计技术的缺失与限制，并提供各种统整性指数与统计量数来检测各项参数与模型特性，例如路径分析的整体效应、间接效应的估计，在 SEM 模型中可以快速有效完成，并提供显著性检验用以检验效应的统计意义，此为结构方程模型最有价值的贡献所在。

二、统合模型的构成

统合模型就是同时包括了测量模型与结构模型的SEM模型，如图 8.1 所示。图 8.1 中，模型a1 与a2 是由三个测量变量所组成的测量模型，模型a3 与a4 是由两个测量变量所组成的测量模型，这四个测量模型所建立的四个潜在变量（或因素）的结构关系形成了一个结构模型，即模型b。四个测量模式的適切性可以分别检验之，也可以以特定的方式组合（例如模型a1 与a2 的潜在变量具有相关），经由CFA分析来检验。另一方面，结构模型的检验则是建立在测量模型的基础上，配合路径分析的策略探讨潜在变量的结构因素关系，F1 与F2 即为外源变量，F3 与F4 为内生变量，外源变量对于内生变量无法解释的部分为残差 D_{F3} 与 D_{F4} ，整个模型即为结构方程模型的统合模型。

图 8.1 可以说是一个完全统合模型，因为结构模型当中的每一个变量都是潜在变量，包括两个外源变量与两个内生变量，潜在变量的估计皆是建立在多重指标所形成的测量模型之上，又称为完全潜在模型（full latent model）。如果结构模型中的某一个或某几个变量是单一指标的测量变量，如图 8.2 的V4 与V7，称为部分潜在模

型 (partial latent model) (Kline, 1998), 也就是说, 模型中夹杂着测量变量与潜在变量。以图 8.2 为例, 其中 V_4 为外源变量, V_7 为内生变量, 两者皆外显的测量变量, 而 D_{V_7} 是外源变量对 V_7 无法解释的残差变异。

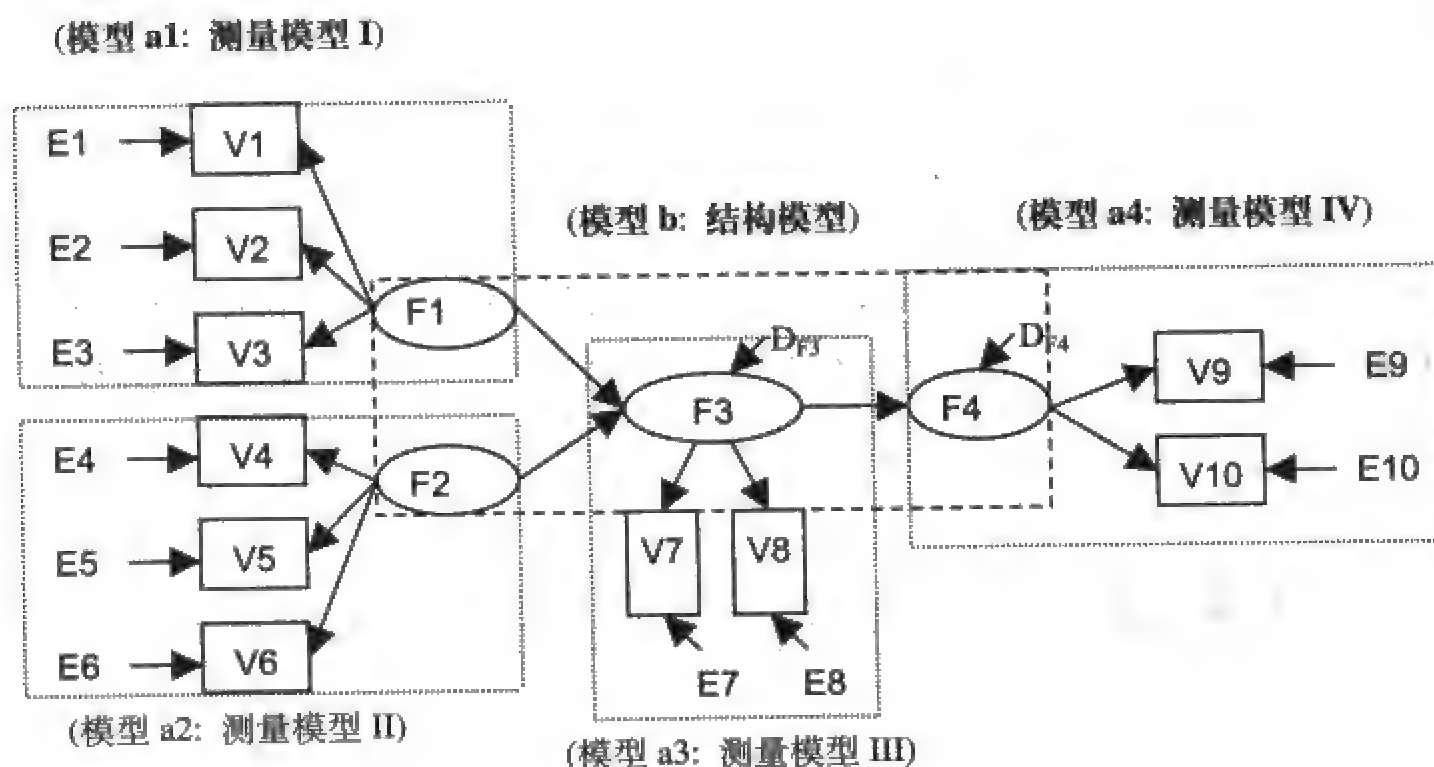


图 8.1 典型的统合模型

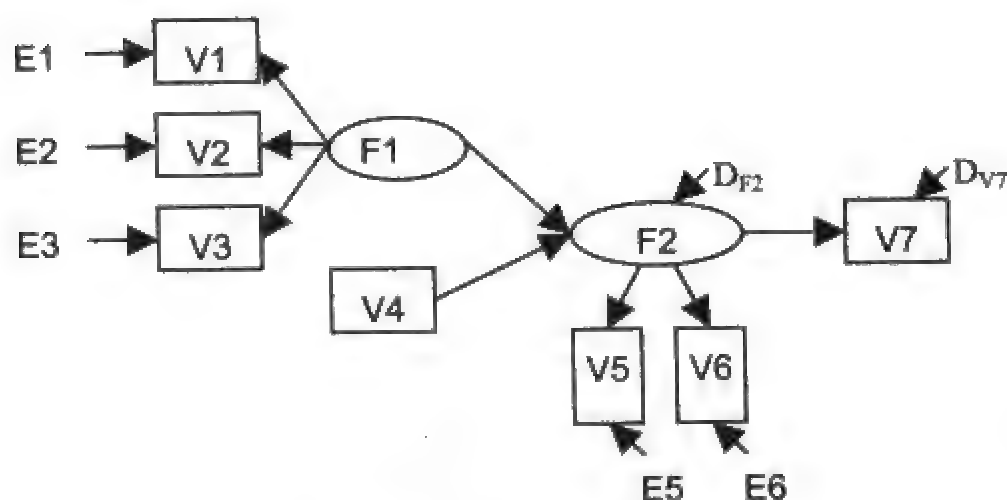


图 8.2 具有两个单一指标测量变量的统合模型

在图 8.1 当中，如果结构模型中的四个潜在变量都以变量组合程序（item parceling）改以单一指标测量变量的形式出现（例如将潜在因素的各指标直接加总得到一个总分），整个模型只剩下外显的测量变量的结构关系，这即是一个典型的回归取向的路径模型。

三、统合模型方程式

（一）测量模型的方程式

图 8.1 当中计有四个独立的测量模型，每一个模型代表了特定的一个潜在变量与两个以上的观察变量之间的组合关系，对于每一个观察变量，可以被视为一个内生变量。它的分数是受到潜在变量（外源变量）的影响以及测量误差（ ε ）的影响，可以一个单独的方程式来表示其间的关系：

$$x = \Lambda_x \xi + \delta \quad (\text{模型 a1 与 a2}) \quad (8.1)$$

$$y = \Lambda_y \eta + \varepsilon \quad (\text{模型 a3 与 a4}) \quad (8.2)$$

上述两个通式当中， ξ 表示潜在外源变量（F1 与 F2）， η 表示潜在内生变量（F3 与 F4）， x 与 y 是各潜在变量所影响的观察变量，在概念上属于内生变量，也就是说，对于每一个个别的观察变量，不论是 x 或 y ，它的分数可以被切割为由共同的潜在变量所影响的部分以及无法被解释的部分（称为测量误差，由 ε 与 δ 表示）。 Λ_x 与 Λ_y 即表示潜在变量各自的因素载荷系数矩阵（matrices of factor loadings）。

以测量模式 I（a1）为例，潜在变量 F1 与三个测量变量的模型关系是由三个联立方程式所组成， Λ_x 表示 F1 潜在变量所影响的三个观察变量 V1、V2、V3 在 F1 上的因素载荷，三个观察变量无法被潜在变量解释的部分为测量误差，分别为 E1、E2、E3。整个模型的组成就是一个典型的单一因素的因素分析模型，其他三个测量模型的方程式是类似的概念。

（二）结构模型的方程式

上述潜在变量与相对应的测量变量的模型关系，显示出测量模型的组成，而结

构模型的组成方程式如下：

$$\eta = B\eta + \Gamma\xi + \zeta \quad (8.3)$$

η 是 $m \times 1$ 个潜在内生变量向量 (vector of endogenous latent variables)，也就是研究者所关心的潜在因变量，即图 8.1 的 F3 与 F4； ξ 是 $k \times 1$ 个潜在外源变量向量 (vector of exogenous latent variables)，是研究者拿来解释或预测因变量的独立变量； B 是 m 个内生变量的 $m \times m$ 回归系数矩阵， Γ 是 m 个内生变量与 k 个外源变量的 $m \times k$ 回归系数矩阵， ζ 为 $m \times 1$ 个内生变量无法被解释的干扰向量 (vector of disturbance)。B 与 Γ 两个矩阵所代表的是方程式的标准化回归系数矩阵，性质等同于传统多元回归的回归系数。多个内生变量的组合关系即为传统的路径分析模型，这些回归系数也就自动成为路径系数，这些系数都是研究者最后所要报告标示出来的重要参数。

四、统合模型的识别

在统合模型当中，由于合并了测量模型与结构模型，牵涉到的参数相当众多且复杂，因此在识别上的处理也就必须格外小心。

首先，在测量模型部分，为了使潜在变量与观察变量之间的关系能够被顺利地估计出来，不论是潜在内生变量或潜在外源变量，各潜在变量之下所属的各观察变量中，挑选其中一个观察变量与潜在变量的参数，也就是因素载荷 (Λ_x 或 Λ_y) 固定为 1.00，以便作为其他参数的参照，并可以让潜在变量的方差自由估计。此时由此一特定的潜在外源变量组成的测量模式就是一个典型的单一因素模型。

一旦测量模型的参照设定完成，整体模型的识别性是否足够可以利用 t 法则计算呢？Bollen (1989) 提出了一个二阶段法则，将统合模型的分析区分为两个步骤。第一个步骤是以因素分析的方式单纯进行测量模型参数估计，第二个步骤则将第一个步骤所建立的潜在因素视为观察变量，以路径分析的模式进行结构模型分析。

第二节 统合模型的分析步骤

基本上，统合模型的执行与先前所介绍的验证性因素分析与路径分析相似，亦即需遵循假设模型的提出、模型识别的计算、参数估计、模型拟合检验到最终解的报告等程序。但是由于统合模式的内容包括测量模型与结构模型两个部分，因此在实际估计时必须考虑这两种模型的执行顺序。如果不区分测量模型与结构模型，而把整个结构方程模型视为一个完整模式，进行一次估计而获得所有的参数估计数，报告一个模式拟合指数评估模型的优劣，称为一阶段策略（one-step modeling）。如果模式拟合不佳，则可进行模式修饰，直到模式拟合良好为止。

相对的，如果一个带有潜在变量的路径分析模型，将测量模型与结构模型分成两个模型进行检测，则称为两阶段策略（two-stage modeling）（James, Mulaik, & Brett, 1982; Anderson & Gerbing, 1989）。第一个阶段确定因素结构的拟合性，第二阶段则是在不改变测量模型的前提下，增加结构模型的设定，并评估结构模型界定之后的拟合性。此种做法是目前广为学者所建议的分析策略。Anderson 与 Gerbing（1988）主张两阶段模型对于潜在变量的效度评估具有重要的价值，因为测量模型的检测可以提供潜在变量的聚敛与区分效度的信息，结构模型则可提供预测效度的证据。

Mulaik 与 Millsap（2000）提出了四阶段的策略，其操作方式是利用一系列的嵌套模型，藉由模式竞争比较策略评估各模型的拟合性。

阶段一：未限制测量模型（unrestricted measurement model）。是一种探索性的测量模式分析，以决定能让理论协方差矩阵与观察协方差矩阵达到最理想拟合下的潜在变量数目。每一个测量变量与每一个潜在变量的因素载荷均加以估计。

阶段二：限制性测量模型（restricted measurement model）。亦即进行研究者所决定的因素结构，确认模式拟合，并决定测量模型的最终参数估计数。不属于某特定潜在变量的因素载荷设定为 0，此举将使此一 CFA 模型会较前一个阶段的模型拟合变差。

阶段三：完整结构模型（structure model）。依据研究者所关心的理论假设关系估计

潜在变量间的参数，无关联的结构参数设定为 0。

阶段四：替代模型竞争比较 (alternative model competition)。依据其他理论观点所提出的结构模型调整结构模型（包括增减结构参数、将结构参数加以设限、或是利用参数的置信区间比较参数的差异），以比较不同结构模型的拟合性，推估最合理的结构模型。

基本上，四阶段策略系从二阶段策略加以扩充，本质并无不同。四阶段的第一阶段之所以需要比较不同数目的潜在变量，多半是因为研究者无法确切掌握因素的结构。在这种情况下，并不适合进行验证性因素分析，而可以改以探索性因素分析先去就因素的可能结构加以讨论。另一方面，第四个阶段实则是第三阶段的结构模型的延伸，如果研究者不关心其他替代模型下的参数关系，并无进一步进行第四阶段的必要。换言之，二阶段策略可以说是结构方程模型分析最根本的操作方法。

多阶段模型的优势是可以获得更丰富的模型拟合数据，藉以厘清影响模式估计的原因为何。Joreskog 和 Sorbom (1983) 与 Bollen (1989) 均曾指出，测量模型的良好拟合提供了后续结构模型的重要基础，因而测量模型必须在结构模式之前进行确认。换言之，如果测量模型不佳，那么结构模型的分析即缺乏合理的测量基础。

从测量模型与结构模型中的参数数目也可以看出，测量模型中所带有的参数远多于结构模型。因此一个良好拟合的测量模型，表示绝大部分的参数（包括因素载荷、测量残差、潜在变量的共变等）都有获得良好合宜的估计值，也确保了后续的结构模型分析的质量。

从参数的性质来看，结构模型的参数多为单方向的回归系数。不论是 $A \rightarrow B$ 或 $B \rightarrow A$ ，其实都是 $A \leftrightarrow B$ 的一种特例。而测量模型系把潜在变量的相关（即 $A \leftrightarrow B$ ）设定为自由估计，因此如果测量模型获得理想拟合，那么结构模型的单方向参数也应会有一定的适切性。Anderson 与 Gerbing (1992) 即主张，测量模型所提供的估计是测量的计量基础，而结构模型所检测的则是理论关系。因此结构模型所看重的并不是测量的合理性，而是参数的方向性背后所代表的理论合理性。

会发生测量模型的参数合理但结构参数不理想的可能性之一是当相关系数过高时，所可能产生的共线性问题。在多元回归当中，多元共线性问题原本就是威胁参数估计与

解释的重要议题 (Cohen, Cohen, West, & Aiken, 2003; Pedhazur, 1997), 而非测量问题。

第三节 变量组合与聚合

结构方程模型分析经常会以变量组合策略 (item parceling) 简化测量模型, 使结构模式得以在比较简化的情形下进行估计 (Bandalos & Finney, 2001)。例如当某个潜在变量有六个观察变量时, 可以将六个变量聚合 (aggregate) 成三个变量 (每两题加在一起或求平均值) 以降低模式的复杂度。聚合层次的变量分数称为组合分数 (parcel scores)。另外, 当研究样本太少 (例如低于 200) 而统合模型很大, 所需估计的参数很多时, 一种变通的做法是将潜在变量改以观察变量处理 (把构成各潜在变量的题目组合成单一变量) 并进行路径分析。

一、变量组合的方法

在结构方程模型分析中, 变量组合并没有一致的方式, 但有一个基本前提是构念具有单维性, 亦即测量同一个构念的测量题目才能够进行组合 (Bandalos & Finney, 2001; Little et al., 2002)。其他的组合方法则各有目的, 研究者必须了解各种方法的特殊性与使用时机, 才能发挥变量组合的优势 (Bandalos & Finney, 2001; Hall et al., 1999)。例如当观察变量的正态性不理想时, 变量组合的方式就应能抵销各题的偏态, 以提高组合后变量的正态性。此外, 也有学者主张必须就题目的内容, 组合题意相近的观察变量, 使得组合后的观察变量拥有可解释的意义 (Comrey, 1970)。

(一) 辐射组合法

较早的一种方法是 Cattell (1974; Cattell & Bursdal, 1975) 所提出的辐射组合法 (radial parceling), 它系利用两阶段的探索性因素分析, 基于观察变量在各萃取因素

当中的因素载荷高低,依序将若干题目组合成组合变量,再进行变量较少的探索性因素分析,确认组合变量是否仍可以得到类似的因素结构。此法适合用于多因素的因素模型(例如个性的测量),但此法费时费力,而且可能会把不同构念的观察题目组合到不当的组群中(Bandalos & Finney, 2001; Barrett & Kline, 1981)。

(二) 随机分布法

最直观的方法是随机分布法(random assignment),其做法是将测量同一个构念的题目以随机方式或系统性方式加以组合(Kishton & Widaman, 1994)。例如一个原有九个随机排列的观察变量的构念,可将第1至3题、第4至6题、第7至9题,依序加总并除以题数3形成三个带有原始测量尺度的新变量,简化成一个带有三个观察变量的测量模式。Little等人(1999; 2002)认为这种策略在多数情况下会比依据题目内容组合会更理想,但是未必会得到效度最好的组合变量。

(三) 项目构念平衡法

为了改善随机法的缺点, Little等人(2002)提出一种更能找到反映同一个构念的一组最相似的组合变量的方法,称为项目构念平衡法(item-to-construct balance approach)。这是利用CFA得到的因素载荷数值,将同一个构念中的题目的载荷,依高低顺序把各题平均分布到各个组合变量中。例如,若要将六个题目组合成三个组合变量,那么因素载荷第1、2、3高的前三个题目依序成为三个组合变量的第一题(定锚题; anchor item),然后载荷次高的另外三题,则以相反的次序(亦即第6、5、4者)归入三个组合变量中成为第二题,使最高载荷的三题与低载荷的三题可以平衡组合在一起。这样使组合后的新变量有一致的变异量。此种方法适用于当研究者想要得到组合后的观察变量有一致的因素载荷时,亦即平行指标(parallel indicators)。

如果研究者同时关心各变量的平均数,则可套用相似的原则,把各题依照截距排列高低,将各题组合成平均数相当的新组合变量。

(四) 同源法

相对于项目构念平衡法所组合的变量是方差或平均数最相当的一组平行测量,另一

种策略则是令所概组合后的因素结构与组合前的结构最相似的同源法 (congeneric parcels) (Fletcher, 2005; Fletcher & Perry, 2007)。具体做法是将最相似的完全标准化因素载荷者组合成一个组合变量, 例如六个题目组合成三个组合变量, 那么因素载荷第 1 或第 2 的最高题目组合成第一个观察变量, 第 3 或第 4 高者组成第二个组合变量, 依此类推。如此一来, 因为组合变量下的观察变量同构型最高, 所形成的组合变量的重要性差异即是原始因素结构内的因素载荷权重排序。因而组合后的新变量进行简化的因素分析时, 可以得到与组合前最相似的因素结构。事实上, 此一做法很类似于辐射组合法以及 Hall 等人(1999) 提出的误差分离法 (isolated uniqueness procedure)。所不同的是它采用的 CFA 分析的数据。Fletcher 与 Perry (2007) 的模拟研究发现此法的效果比项目构念平衡法为优。但是, 此法的适应性还有赖更多的研究证实。

二、变量组合的优缺点

如果第一阶段所获得的测量模型的拟合性理想时, 后续利用组合分数进行组合分数的统合分析会有诸多优点。除了模型可大幅简化、提高模型拟合度、获得较理想的估计解之外 (Marsh, Hau, Balla, & Grayson, 1998; Rogers & Schmitt, 2004; Yuan, Bentler, & Kano, 1997), 也可以让测量模式得到较佳的测量信度 (Cattell & Burdsal, 1975; Kishton & Widaman, 1994)、较高的变量解释力或共同性 (communality) (Little, Cunningham, Shahar, & Widaman, 2002)、观察变量的尺度更具有等距性与正态性 (Bagozzi & Heatherton, 1994)、更理想的检验效力 (每题的样本数与题数之比值较高) (Bagozzi & Edwards, 1998; Gottfried, Fleming & Gottfried, 1994; Vandenberg & Scarpello, 1991)、避免特殊题目的干扰 (Chapman & Tunmer, 1995)。

Bandalos 与 Finney (2001) 说明了组合分数之所以会有较佳的拟合, 主要是因为当利用组合分数时, 观察变量矩阵中的元素数目会大幅度降低, 同时也让观察矩阵与导出矩阵的差异 (残差) 减少, 进而大幅提升模型拟合的数据。另外, 各潜在变量的信度会提高, 是因为潜在变量对观察变量的解释力提高 (共同性提高), 测量误差的比例也就随之下降。因此, 根据此一原理, 在进行两阶段的统合模型估计时, 如果测量模型并不十分理想时, 可以在第一阶段尝试将测量模型简化, 以利第二阶段的结构模型的估计。

Bandalos 与 Finney (2001) 整理了七个期刊当中的论文发现, 317 篇当中有 62 篇 (占 20%) 涉及了组合变量的运用。

组合变量的处理虽然有诸多优点, 但是也有一些限制。最重要的一点是研究者不应为简化而简化, 而需提出简化的理由与统计量据上的支持。例如, 能够加以组合的题目必定具有相同的量尺 (例如都是二分变量或是同点数的 Likert 量尺)。此外, 能够组合成简化题目的一组观察变量, 应具有单维性 (uni-dimensionality), 意即它们是测量同一个构念的变量, 受到同一个潜在变量的影响。如果一群观察变量的相关并不高, 或是其性质是在测量不同的构念, 且具有多维性 (multidimensionality) 时, 即不宜组合成组合变量 (Bandalos, 2002; Bandalos & Finney, 2001; Hall, Snell, & Foust, 1999; Kim & Hagtvet, 2003; Little et al., 2002)。因为观察变量之间的关系可能非常复杂, 如果径自组合变量, 可能会遮蔽了有意义的因素结构, 造成不当的测量模型。

一般在进行变量组合之前, 必须经过验证性因素分析的检验, 以确保构念的单维性。如果观察变量的因素载荷越高, 表示单维性越稳固, 越适合进行题目组合。进一步地, 在高阶验证性因素分析当中, 如果第一阶潜在变量能够形成更高阶的二阶潜在因素, 表示这些低阶因素具有高阶的单维性, 因此也可以利用组合变量策略, 把第一阶的潜在变量改以测量变量处理, 用来定义高阶的潜在变量。

Meade 与 Kroustalis (2006) 引入了测量恒等性 (measurement invariance) 的概念, 强调组合变量的使用必须在这些变量在不同情境 (例如前后测) 或不同团体 (例如性别) 下具有测量恒等性的前提下, 才能确保组合过程不至于产生团体偏误。在恒等性的评估时, 最重要的是去检测因素载荷。因为变量的组合直接影响变量的方差与共变量的数值, 所以组合后的观察变量的因素载荷在跨样本间都应具有恒等性。如果研究者关心潜在变量的平均数, 那么组合变量的截距的恒等性就必须纳入考虑。Meade 与 Kroustalis (2006) 以仿真数据获得的研究建议是, 为了能够让组合变量的恒等性得以确保, 研究者应在组合之前逐一检视个别观察变量的恒等性, 找出导致不恒等的变量, 才比较有可能得到理想的恒等性结论。其他有关变量组合的问题可以参考相关的文献 (Bandalos, 2002; Bandalos & Finney, 2001; Landis, Beal, & Tesluk, 2000; Little et al., 2002; Nasser & Wisenbaker, 2003; Rogers & Schmitt, 2004)。

第四节 LISREL 的统合模型分析

前一个范例示范了潜在变量路径分析的典型应用。通常一个统合模型分析当中包含有多个潜在变量。这些潜在变量除了作为外源变量与内生变量，有时还扮演着中介变量的角色，也就是介于两个潜在变量之间；即作为被他人影响的因变量，也作为影响他人的自变量。其他变量对于因变量的影响，可以透过该中介变量间接影响因变量。此时，我们可以称之为中介模型的潜在变量路径分析。

中介模型的分析，其基本原理与前面范例五的做法类似，然而更重视间接效应的分析。在模型界定上，则必须涉及测量模型与结构模型的设定。因此，在分析的复杂度上，会更甚于单纯的路径分析或验证性因素分析。

本范例将示范中介模型 SEM 分析的具体做法，并使用实际研究数据库，使读者可以明了真实研究数据的分析情形。所使用的资料为 250 位基层教师的创意教学行为调查数据。调查数据包括了教师的创造人格特质、创意教学自我效能感、组织社会化与教学创新行为。其中创造人格特质的测量系透过叶玉珠、吴静吉、郑英耀等人（2000）所编制的《创造性格量表》摘要题本，包括三个测量变量“多角推理”、“兴趣广泛”、“乐在工作”，创意教学自我效能感的测量系采林碧芳和邱皓政（2003）所编制的《创意教学自我效能感量表》，共计有四个指标：“自我肯定”、“自我防卫”、“社会支持寻求”、“外在压力抗衡”。

组织社会化程度的测量则取自郭美凰（1996）的量表，经重新分析后，得到“组织融入”与“工作熟练度”两个指标；教师教学创新行为则是采吴静吉、王文中、郭俊贤、陈淑惠与李慧贤（1996）所编制的教师教学创新行为量表，共计九题，经加总后得到一个总分，分数越高，代表教师的教学创新行为越多。以下，我们将利用 SEM 的统合模型分析技术进行潜在变量的中介模型估计程序的分析。

一、假设模型

本范例所采用的数据库,旨在探讨教师的创意教学行为是否受到教师自身对于执行创意教学的胜任感的自我评价影响(也就是创造行为的自我效能感)。也就是说,当教师的创意教学自我效能感越强,越有可能展现出创意教学的行为。然而,教师创意教学的自我效能感的高低,可能受到教师个人性格特质与组织生活经验(组织社会化程度)的影响,因此,研究假设可写为:

1. 教师创意教学自我效能越高,越能够表现出创意教学行为。
2. 教师个人创造性格越强、组织社会化程度越高,则创意教学自我效能越高。
3. 教师个人创造性格与组织社会化程度两项特质,会透过自我效能的中介作用,间接影响教师的创意教学行为。

根据上述假设,以及前面所提及的各变量的测量指标,我们可以得到图 8.3 的概念图。

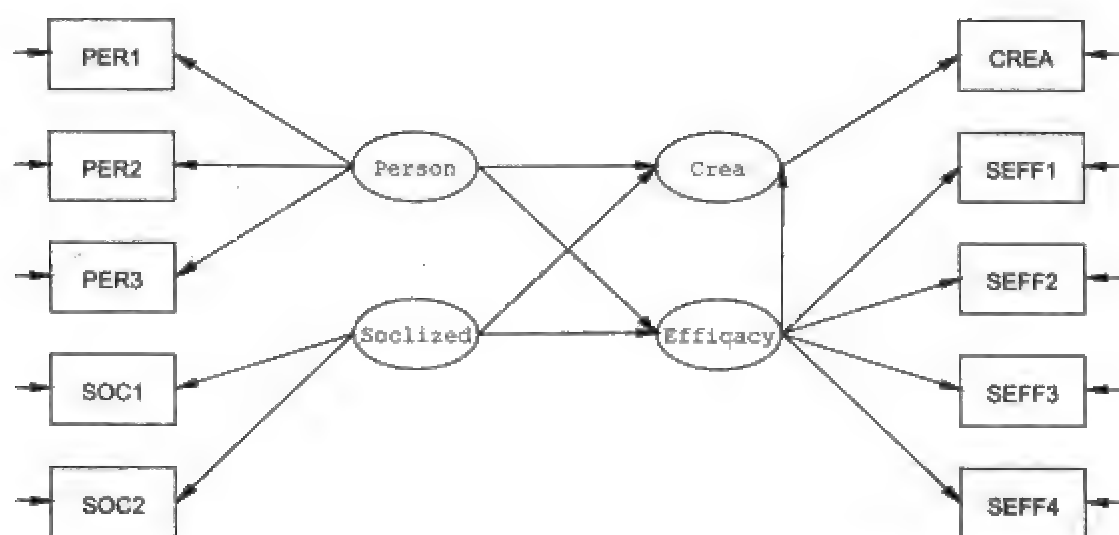


图 8.3 教师创意教学行为研究的 SEM 统合模型的概念图示

二、模型界定

整个模型当中,共有 10 个测量变量,测量资料数为 $(10 \times 11)/2 = 55$ (DP=55)。整个

模型的设定条件如下：

1. 模型中有 5 个外源测量变量 (x_1 至 x_5)，5 个内生测量变量 (y_1 至 y_5)。
2. 模型中有 2 个外源潜在变量 (ξ_1 、 ξ_2) 与 2 个内生潜在变量 (η_1 、 η_2)。
3. 模型中有 5 个外源测量残差 (δ_1 至 δ_5)，5 个内生测量残差 (ε_1 至 ε_5)、2 个解释残差 (ζ_1 、 ζ_2)；其变异量被自由估计。
4. 内生潜在变量被外源潜在变量解释，Beta矩阵有 1 个结构参数 (β_1)。Gamma矩阵有 4 个结构参数 (γ_{11} 、 γ_{12} 、 γ_{21} 、 γ_{22})。
5. 每一个测量变量仅受单一个潜在变量影响 (单维假设)，故产生 5 个外源测量变量因素载荷参数 (λ_{x1} 至 λ_{x5}) 与 5 个内生测量变量因素载荷参数 (λ_{y1} 至 λ_{y5})。
6. 为了使潜在变量的量尺得以确立，各潜在变量的第一个因素载荷被设定为 1，共有 4 个因素载荷被设定为 1。

三、参数估计与分析

为了示范两阶段法，此步骤中首先将进行一个带有四个具有相关因素的 CFA，然后再进行第二阶段的结构模型分析。本范例所输入的数据为协方差矩阵，内容如下 (Ch8.cov)：

```

17.711
1.843 0.404
0.801 0.146 0.374
1.243 0.227 0.110 0.589
1.692 0.226 0.115 0.208 0.393
7.518 1.074 0.301 0.531 0.741 7.565
7.585 1.062 0.538 0.609 0.806 5.202 7.046
6.499 0.906 0.388 0.434 0.664 4.530 4.626 6.335
3.020 0.369 0.237 0.322 0.196 1.947 1.524 1.649 4.482
2.663 0.318 0.307 0.335 0.326 2.092 1.824 1.662 2.335 2.982

```

(一) 阶段一：CFA

1. 分析语法

CFA 的 LISREL 语法文件与 SIMPLIS 语法文件列于表 8.1, 数据文件读入的变量排列顺序如下: 创意教学经验 (CREA), 创意教学自我效能的“自我肯定效能”、“自我防卫效能”、“支持寻求效能”、“压力抗衡效能” (SEFF1-SEFF4), 创造人格量表的“独特思考”、“广泛兴趣”、“乐在工作” (PER1-PER3), 以及组织社会化的“组织融入”与“工作熟悉” (SOC1-SOC2)。

表 8.1 统合模型阶段一的 LISREL 与 SIMPLIS 语法(ch8a.ls8 与 ch8a.spl)

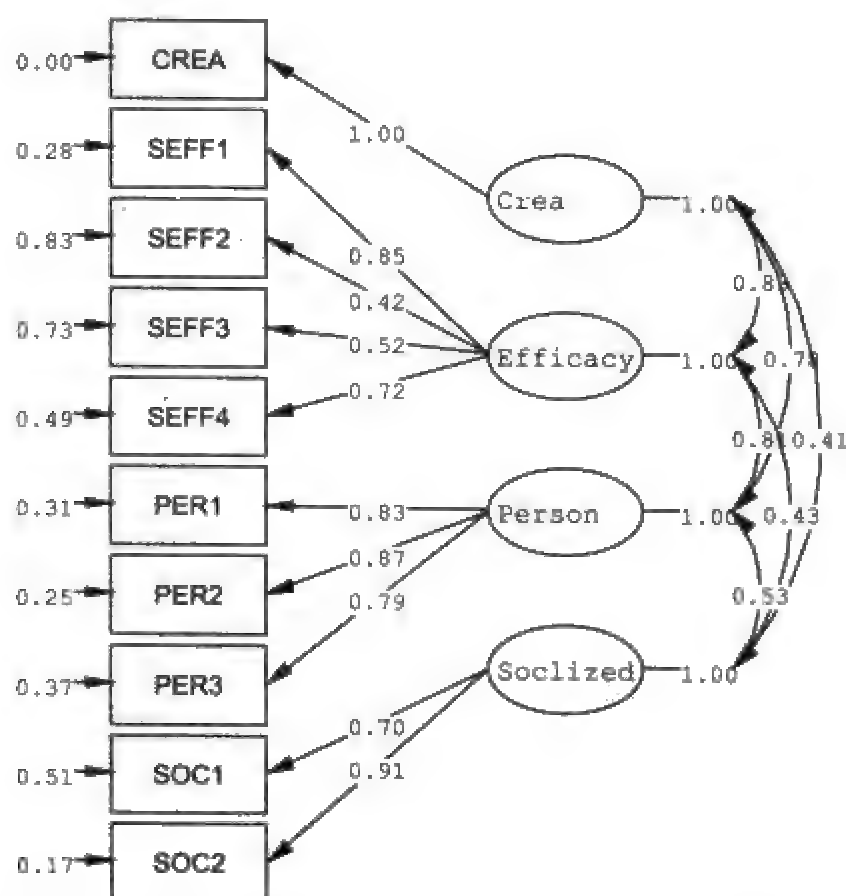
<i>LISREL</i>	
1	Ch8 SEM Stage 1 CFA using LISREL syntax
2	DATA NI=10 NO=250
3	CM SY FI=Ch8.cov
4	LA; CREA SEFF1 SEFF2 SEFF3 SEFF4 PER1 PER2 PER3 SOC1 SOC2
5	MO NX=10 NK=4
6	LK; Crea Efficacy Person Soclized
7	FR LX (3,2) LX(4,2) LX(5,2) LX(7,3) LX(8,3) LX(10,4)
8	FI TE 1
9	VA 1 LX 1 1 LX 2 2 LX 6 3 LX 9 4
10	PD
11	OUTPUT SE TV RS SS SC MI
<i>SIMPLIS</i>	
1	Ch8 SEM Stage 1 CFA using SIMPLIS syntax
2	Observed variables: CREA SEFF1 SEFF2 SEFF3 SEFF4 PER1 PER2 PER3 SOC1 SOC2
3	Covariance matrix from file Ch8.cov
4	Sample Size = 250
5	Latent Variables Creat Efficacy Person Soclized
6	Relationships
7	CREA = Creat
8	SEFF1-SEFF4 = Efficacy
9	PER1-PER3= Person
10	SOC1 -SOC2 =Soclized
11	Set the error variance of CREA to 0
12	Path Diagram
13	End of Problem

值得注意的是，由于创意教学经验（CREA）只有单一观察变量，因此误差设为 0，在 LISREL 语法中有第 8 行(FI TE 1 1)为 0，在 SIMPLIS 语法中则有 Set the error variance of CREA to 0。

至于创意教学经验（CREA）的唯一观察变量的因素载荷设为 1.00，在 LISREL 语法中的第 9 行有 VA 1 LY 1 1，在 SIMPLIS 则不必加以设定（因为预设第一条参数即为 1.0 参数）。

2. 分析结果

测量模型的检测结果列于图 8.4。模式的拟合情形大致良好，WLS $\chi^2=72.79$ ($df=30$, $p=0.00$)，卡方自由度比为 2.426，亦接近 2 的惯用门槛。RMSEA 值为.076；NNFI=.96、CFI=.98，均达.95 标准。因素载荷也大致良好，因此继续进行第二阶段的结构模型的估计。



Chi-Square=72.79, df=30, P-value=0.00002, RMSEA=0.076

图 8.4 统合模型分析阶段——CFA 估计结果图

表 8.2 统合模型阶段二的 LISREL 与 SIMPLIS 语法(ch8b.ls8 与 ch8b.spl)

LISREL

```

1      Ch8 SEM Stage 2 Structure Model Analysis using LISREL syntax
2      DATA    NI=10 NO=250
3      CM SY    FI=Ch8.cov; LA FI=Ch8.lab
4      MODEL    NY=5 NE=2 NK=2 NX=5 PS=DI GA=FU BE=FU
5      LE; Crea Efficacy
6      LK; Person Soclized
7      FR LY(3,2) LY(4,2) LY(5,2) LX(2,1) LX(3,1) LX(5,2) BE (1,2)
8      FIX TE 1
9      VALUE 1 LY 1 1 LY 2 2 LX 1 1 LX 4 2
10     PD
11     OUTPUT SE TV RS FS EF SS SC MI

```

SIMPLIS

```

1      Ch8 SEM Stage 2 Structure Model Analysis using SIMPLIS syntax
2      Observed variables: CREA SEFF1 SEFF2 SEFF3 SEFF4 PER1 PER2 PER3 SOC1 SOC2
3      Covariance matrix from file Ch8.cov
4      Sample Size = 250
5      Latent Variables  Creat Efficacy Person Soclized
6      Relationships
7      CREAT = Crea
8      SEFF1-SEFF4 =Efficacy
9      PER1-PER3=Person
10     SOC1-SOC2 =Soclized
11     Crea = Efficacy
12     Crea = Person Soclized
13     Efficacy = Person Soclized
14     Set the error variance of CREAT to 0
15     Path Diagram
16     End of Problem

```

（二）阶段二：结构模型分析

1. 分析语法

表 8.2 中 LISREL 语法共有 11 行指令，DATA 指令说明变量数目为 10，样本数目 250。CM SY 表示数据以对称的协方差矩阵形式读入，放置于 Ch8.cov 档案中，10 个观察变量的标签放置于外部档案 Ch8.lab，以 LA FI=Ch8.lab 读入。

第 4 行的 Model 指令说明模型的条件：内生观察变量（NY=5）与外源观察变量（NX=5）数目各为 5，内生潜在变量（NE=2）与外源潜在变量（NK=2）数目各为 2，内生变量残差矩阵为对角线矩阵并自由估计（PS=DI），结构系数矩阵为完全矩阵并先加以估计（GA=FU;BE=FU）。

第 7 行开始的参数设定指令中，FR 说明了需要估计的各因素载荷。第 8 行的 FI 指令则指出第一个内生测量变量的残差设定为 0（TE 1），表示第一个内生潜在变量（创意教学行为）是由单一指标所构成，创意行为潜在变量实质上是一个测量变量。第 9 行则指定设定为 1 的参数，以使每一个潜在变量得以量尺化。

LIMPLIS 语法从第 6 行开始，说明测量模型与结构模型的设定方式，测量模型由第 7 至第 10 行，结构模型由第 11 到第 13 行。以第 13 行为例，说明自我效能有两个预测变量 Efficacy = Person Soclized，该行指令也可以写作：Person Soclized-> Efficacy，结果不变。

2. 分析结果

根据上述的模型界定，以 LISREL 软件分析的结果（ex6.out）如下：

```
Number of Input Variables 10
Number of Y - Variables   5
Number of X - Variables   5
Number of ETA - Variables  2
Number of KSI - Variables  2
Number of Observations    250
```

总计读入 10 个变量，内生与外源测量变量各为 5，内生与外源潜在变量各为 2，样本数为 250。紧接着列出所读入的协方差矩阵与平均数数据，在此予以省略。

(1) 参数估计结果

由参数设定的列表可以看出各参数的估计情形。Lambda-Y 与 Lambda-X 参数为测量模式内的因素载荷, 编号为 1 至 3 与 4 至 6; Beta 系数 (7) 系指自我效能对创意行为的回归系数; Gamma 矩阵估计所有的四个参数 (8 至 11); 其他参数均为残差参数 (Phi、Theta-Eps 与 Theta-Delta), 编号 12 至 25。共有 25 个被估计参数。

Parameter Specifications

LAMBDA-Y			外源变量测量模型设定值。
	Crea	Efficacy	
CREAT	0	0	
SEFF1	0	0	
SEFF2	0	1	
SEFF3	0	2	
SEFF4	0	3	
LAMBDA-X			内生变量测量模型设定值。
	Person	Soclized	
PER1	0	0	
PER2	4	0	
PER3	5	0	
SOC1	0	0	
SOC2	0	6	
BETA			结构模型设定值 (内生潜在变量之间的关系)。
	Crea	Efficacy	
Crea	0	7	
Efficacy	0	0	
GAMMA			结构模型设定值 (外源到内生潜在变量之间的关系)。
	Person	Soclized	
Crea	8	9	
Efficacy	10	11	
PHI			外源潜在变量的方差与协方差。
	Person	Soclized	
Person	12		
Soclized	13	14	
PSI			内生潜在变量解释后残差方差。
	Crea	Efficacy	
	15	16	
			内生观察变量残差。第一个题目的残差设定为 0。

THETA-EPS

CREAT	SEFF1	SEFF2	SEFF3	SEFF4
-----	-----	-----	-----	-----
0	17	18	19	20

THETA-DELTA

PER1	PER2	PER3	SOC1	SOC2
-----	-----	-----	-----	-----
21	22	23	24	25

外源观察变量残差。

参数估计以最大概似法进行估计，耗费 12 次迭代。结果发现各测量模型的因素载荷均达显著水平，表示测量模型良好（注 1）。

Number of Iterations = 12

LISREL Estimates (Maximum Likelihood)

LAMBDA-Y

	Crea	Efficacy
-----	-----	-----
CREAT	1.00	- -
SEFF1	- -	1.00
SEFF2	- -	0.47 (0.07) 6.41
SEFF3	- -	0.74 (0.09) 8.10
SEFF4	- -	0.83 (0.07) 11.96

注 1：各因素载荷量估计数均达显著水平，表示测量模型理想。

LAMBDA-X

	Person	Socialized
-----	-----	-----
PER1	1.00	- -
PER2	1.01 (0.06) 15.85	- -
PER3	0.87 (0.06) 14.10	- -

SOC1	- -	1.00
SOC2	- -	1.06 (0.15) 7.13

Beta 系数亦达显著水平，表示自我效能感可以有效预测创意教学行为（估计值为 4.36，标准误为.81， $t=5.4$ ， $p<.01$ ）（注 2）。但组织社会化对于两个内生潜在变量预测值（Gamma 系数）则未达显著水平（注 3）。

BETA			
	Crea	Efficacy	
Crea	- -	4.36 (0.81) 5.40	注 2：回归系数 Beta 估计数为 4.36， t 值为 5.4，达显著水平。
Efficacy	- -	-	
GAMMA			
	Person	Soclized	
Crea	0.62 (0.19) 3.26	-0.01 (0.15) -0.07	注 3：回归系数 Gamma 有两个未达显著水平。表示该两参数没有统计意义。
Efficacy	0.19 (0.02) 9.92	0.00 (0.03) 0.05	

外源潜在变量间的共变与相关估计结果显示，两个外源潜在变量间的方差估计值达显著水平，而两个外源潜在变量间的共变估计值亦达显著水平（估计值为 1.79，标准误为.35， $t=5.10$ ， $p<.01$ ）（注 4）。

Covariance Matrix of ETA and KSI				
	Crea	Efficacy	Person	Soclized
Crea	17.71			
Efficacy	1.87	0.29		
Person	7.50	0.99	5.19	
Soclized	2.59	0.34	1.79	2.19

PHI

	Person	Soclized
Person	5.19 (0.67) 7.74	
Soclized	1.79 (0.35) 5.10	2.19 (0.44) 4.98

注 4：外源潜在变量间的共变与相关估计值，5.19 与 2.19 分别为两个外源潜在变量间的方差估计值，1.79 则为两个外源潜在变量间的共变估计值，均达显著水平。

PSI

Note: This matrix is diagonal.

Crea	Efficacy
4.96 (0.60) 8.22	0.10 (0.02) 4.94

注 5：结构模型对于潜在变量的解释力，类似于 R^2 的功能。表示创意行为可以被解释 72% 的变异量，自我效能可以被解释 65% 的变异量。

Squared Multiple Correlations for Structural Equations

Crea	Efficacy
0.72	0.65

Squared Multiple Correlations for Y - Variables

CREAT	SEFF1	SEFF2	SEFF3	SEFF4
1.00	0.72	0.17	0.27	0.51

THETA-DELTA

PER1	PER2	PER3	SOC1	SOC2
2.37 (0.28) 8.34	1.76 (0.24) 7.29	2.36 (0.26) 9.01	2.29 (0.34) 6.70	0.50 (0.31) 1.59

Squared Multiple Correlations for X - Variables

PER1	PER2	PER3	SOC1	SOC2
0.69	0.75	0.63	0.49	0.83

注 6：外源测量模型的残差与解释力分析。数据显示组织社会化变量对工作熟悉变量的解释力较高。

(2) 模型拟合度分析

拟合度分析数据显示拟合度尚理想, WLS 卡方值为 72.78 ($df=30$, $p=0.00$), 卡方自由度比为 2.426, 亦接近 2 的惯用门槛。*RMSEA* 值为 .076, *NNFI* = .96、*CFI* = .98, 均达 .95 标准; 尽管如此, 此模型仍有修正空间。

Goodness of Fit Statistics

Degrees of Freedom = 30
 Minimum Fit Function Chi-Square = 80.64 ($P = 0.00$)
 Normal Theory Weighted Least Squares Chi-Square = 72.79 ($P = 0.00$)
 Estimated Non-centrality Parameter (NCP) = 42.79
 90 Percent Confidence Interval for NCP = (21.54 ; 71.74)

Minimum Fit Function Value = 0.32
 Population Discrepancy Function Value (F0) = 0.17
 90 Percent Confidence Interval for F0 = (0.087 ; 0.29)
 Root Mean Square Error of Approximation (RMSEA) = 0.076
 90 Percent Confidence Interval for RMSEA = (0.054 ; 0.098)
 P-Value for Test of Close Fit (RMSEA < 0.05) = 0.029

Expected Cross-Validation Index (ECVI) = 0.49
 90 Percent Confidence Interval for ECVI = (0.41 ; 0.61)
 ECVI for Saturated Model = 0.44
 ECVI for Independence Model = 8.93

Normed Fit Index (NFI) = 0.96
 Non-Normed Fit Index (NNFI) = 0.96
 Parsimony Normed Fit Index (PNFI) = 0.64
 Comparative Fit Index (CFI) = 0.98
 Incremental Fit Index (IFI) = 0.98
 Relative Fit Index (RFI) = 0.95

Critical N (CN) = 158.14

Root Mean Square Residual (RMR) = 0.10
 Standardized RMR = 0.042
 Goodness of Fit Index (GFI) = 0.94
 Adjusted Goodness of Fit Index (AGFI) = 0.90
 Parsimony Goodness of Fit Index (PGFI) = 0.52

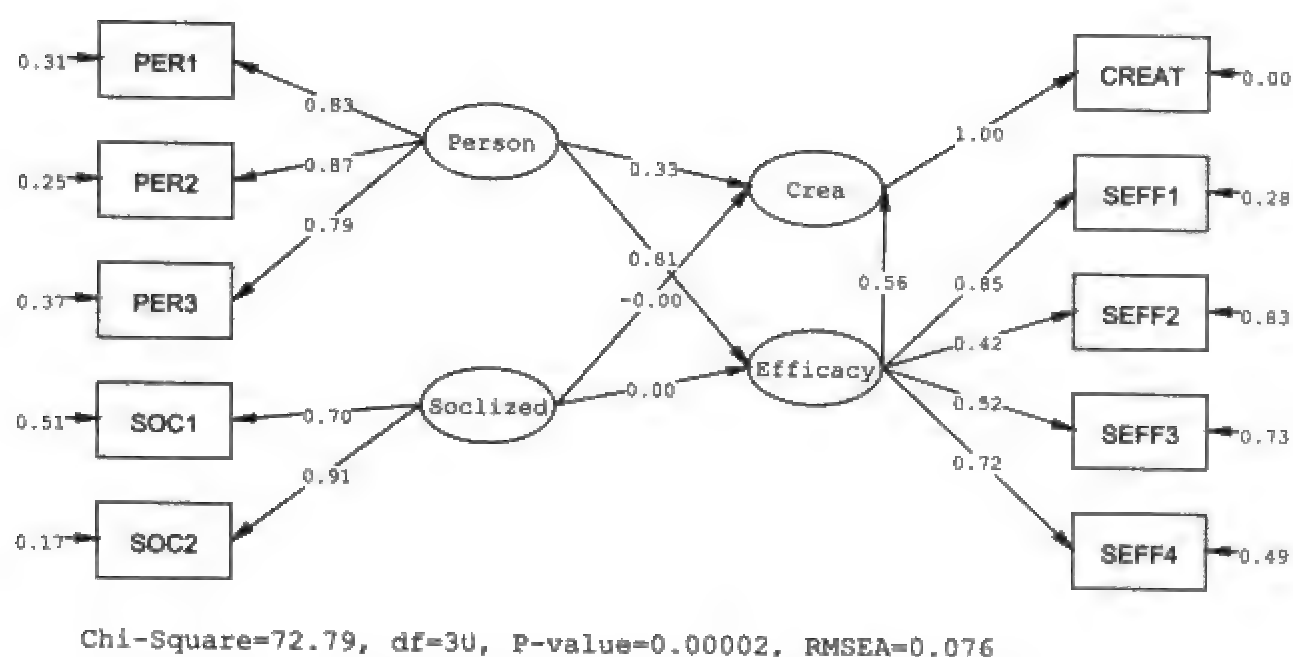


图 8.5 统合模型阶段二结构模型估计结果图

四、模型的修饰

延续前一个步骤，我们可以利用修饰指数检查可以进行增补的参数。LISREL 以提供路径图的形式来说明，如图 8.6 所示。

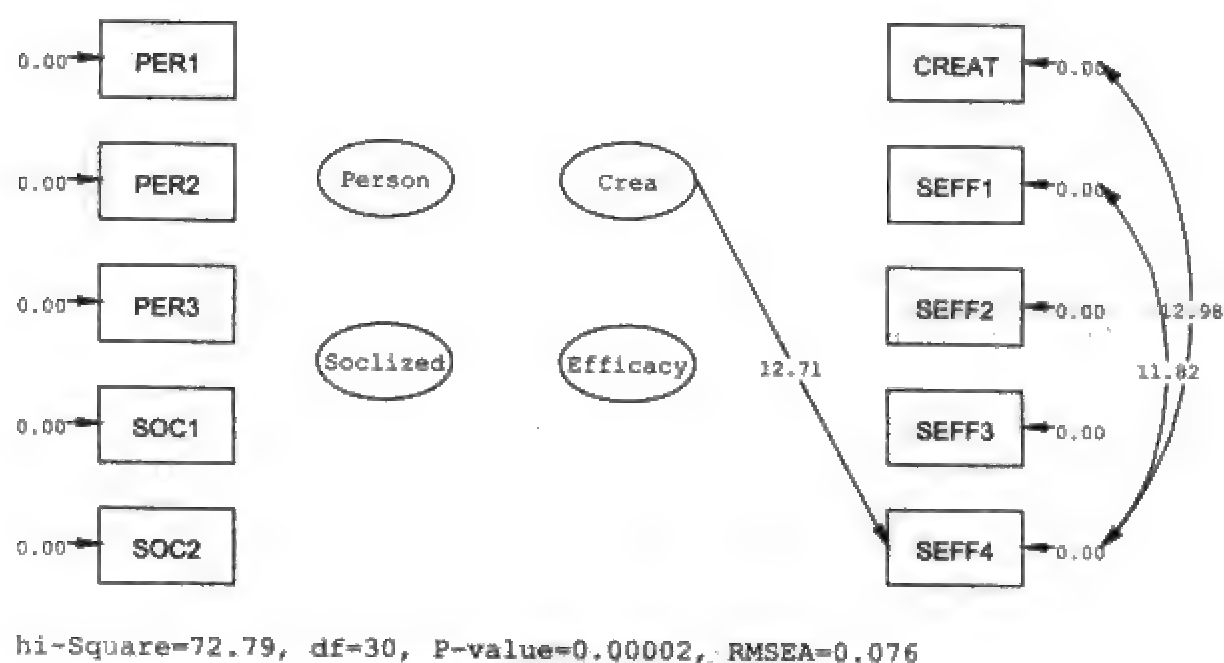


图 8.6 统合模型修饰指数图

图 8.6 显示有三条路径被建议纳入模型中,其中包括一条由潜在内生变量“创意教学行为”(Crea)影响自我效能的压力抗衡变量(SEFF4)的因素载荷参数($MI=12.71$)。另外两个参数则为测量残差。一般而言,参数的增补必须优先考虑理论的適切性,然后再以一般参数优先,残差的相关假设应小心为之。

以本范例来说 Crea 影响 SEFF4 的因素载荷参数有其逻辑上的合理性,表示创造教学行为这一个潜在变量影响了自我效能的抗压信念。当教师表现出创意行为之后,可能助长其抗压信念,因此,可以建议优先纳入修补参数当中。

但是值得注意的是,由 Crea→SEFF4 的这条参数属于测量模式的参数,理应在第一阶段 CFA 分析时进行模型修饰。到了第二阶段的结构模式估计时,即不应再去更动测量模式的参数,在此仍进行修正处理,仅作为示范的目的。

(一) 模型修饰后的新语法

为了增加上述新增因素载荷,在表 8.1 的 LISREL 语法中,第 7 行的 FREE 指令增加一条 LY 5 1 的参数,表示第一个潜在变量(创造教学行为)对于第五个测量变量的因素载荷需加以估计。至于另外两个残差相关是否列入修饰,则可等待 LY 5 1 估计完成后,再以 MI 指数评估之。因为纳入了 LY 5 1 参数后,其他参数的修正建议指数也可能改变。

值得注意的是,在参数估计时,亦发现组织社会化(Soclized)因素对于两个内生变量的解释力未达显著水平,因此该两条 Gamma 参数也可以予以取消,一并纳入修正。但为了简化修饰的内容,本范例先行处理 LY 5 1 的修正(语法文件为 Ch8mo1.ls8)。

(二) 模型修饰后的结果

新增的因素载荷参数,估计的结果发现该参数的因素载荷为.39,达显著水平($t=4.61, p<.001$)。整体模式拟合度也稍有改善, RMSEA 指数为.071,卡方值为 65.15,卡方值改变量为 7.64 ($df=1, p<.01$),表示该参数的增加有统计意义(结果档为 Ch8mo1.out)。新增的参数的统计显著性被确认之后,我们再检查 MI 指数,发现仍有一个残差相关得以增补(SEFF3 与 SEFF4 的残差相关, $MI=8.74$)。因此,若在模型中纳入该参数后,重新加以估计,语法文件为 Ch8mo2.ls8,发现该相关系数为.13

($t=2.86, p<.01$), 具有统计意义。整体模式拟合度也更为改善, RMSEA 指数为.063, 卡方值为 55.59, 卡方值改变量为 9.56 ($df=1, p<.01$), 表示该参数的增加有统计意义 (结果档为 Ch8mo2.out)。再以 MI 指数检查, 已无任何建议参数, 所得到的最终模型可以作为分析终解。用户也可以利用 LISREL 的路径图指令 (PD), 获得图 8.7 的标准化系数的路径图。

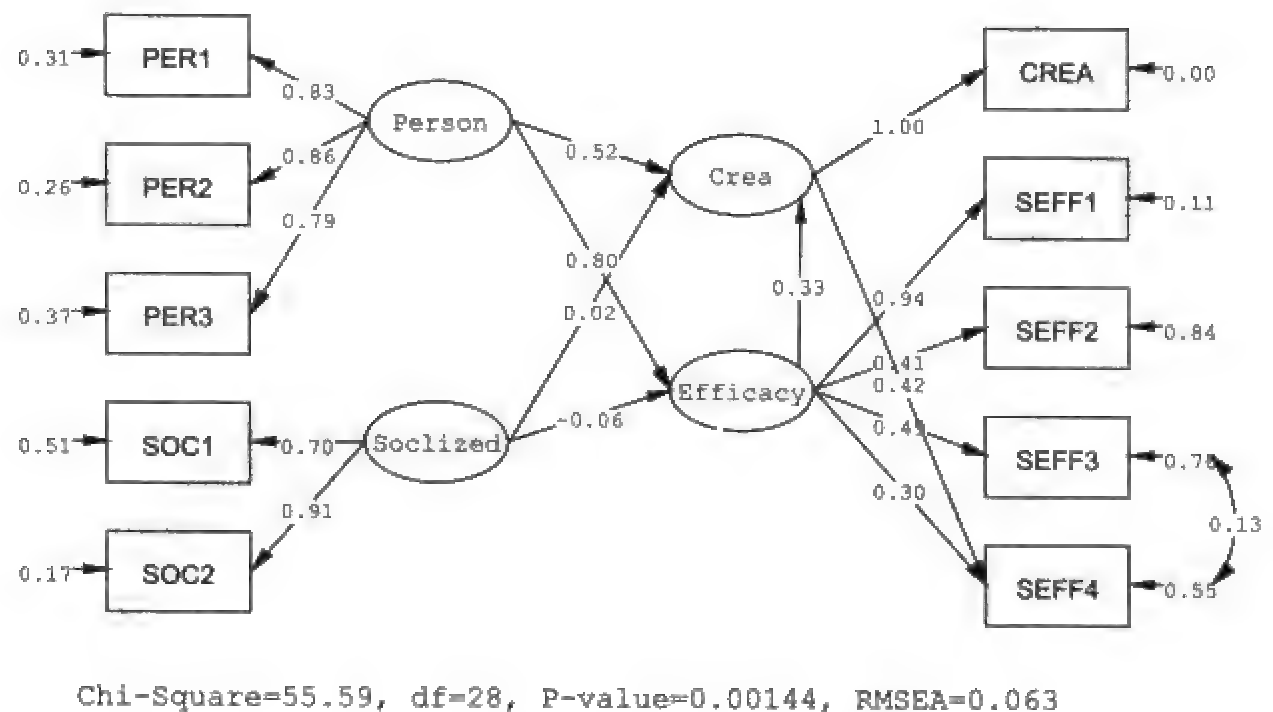


图 8.7 修正后统合模型最终解路径图

五、完成 SEM 分析

(一) 标准化参数估计数

中介模型的潜在变量路径分析所产生的数据中, 兼含测量模型 (注 7) 与结构模型 (注 8) 的数据。而所有参数的完全标准化解, 除了以路径图来表示, 也可以由报表整理而得 (Ch8mo2.out)。

Completely Standardized Solution

LAMBDA-Y		
	Crea	Efficacy
CREA	1.00	-
SEFF1	-	0.94
SEFF2	-	0.41
SEFF3	-	0.49
SEFF4	0.42	0.30

注 7: 各测量模型的标准化系数, 也就是传统的因素载荷量。

LAMBDA-X		
	Person	Soclized
PER1	0.83	-
PER2	0.86	-
PER3	0.79	-
SOC1	-	0.70
SOC2	-	0.91

BETA		
	Crea	Efficacy
Crea	-	0.33
Efficacy	-	-

注 8: 结构模型的标准化系数, 也就是 0 传统的回归系数。社会化因素的回归系数没有统计意义。

GAMMA		
	Person	Soclized
Crea	0.52	0.02
Efficacy	0.80	-0.06

Correlation Matrix of ETA and KSI

	Crea	Efficacy	Person	Soclized
Crea	1.00			
Efficacy	0.73	1.00		
Person	0.78	0.77	1.00	
Soclized	0.42	0.37	0.53	1.00

PSI
Note: This matrix is diagonal.

	Crea	Efficacy
	0.34	0.41

THETA-EPS

	CREA	SEFF1	SEFF2	SEFF3	SEFF4
CREA	- -				
SEFF1	- -	0.11			
SEFF2	- -	- -	0.84		
SEFF3	- -	- -	- -	0.76	
SEFF4	- -	- -	- -	0.13	0.55

THETA-DELTA

	PER1	PER2	PER3	SOC1	SOC2
	0.31	0.26	0.37	0.51	0.17

Regression Matrix ETA on KSI (Standardized)

	Person	Soclized
Crea	0.78	0.00
Efficacy	0.80	-0.06

(二) 直接与间接效应

由报表数据可知，与组织社会化有关的直接与间接效应均未达显著水平（注 9 与注 10），显示组织社会化变量作为一个代表外在环境影响的变量，在整个模型中解释力甚低。

Total and Indirect Effects

Total Effects of KSI on ETA

	Person	Soclized
Crea	1.44 (0.12) 11.66	-0.01 (0.17) -0.03
Efficacy	0.21 (0.02) 10.66	-0.02 (0.03) -0.85

注 9：结构模型中，与组织社会化有关的所有整体效应均未达显著。显示组织社会化变量的预测效果微弱。

Indirect Effects of KSI on ETA

	Person	Soclized
Crea	0.48 (0.14) 3.39	-0.05 (0.06) -0.83

Total Effects of ETA on ETA

	Crea	Efficacy
	-----	-----
Crea	- -	2.30 (0.67) 3.42

Total Effects of ETA on Y

	Crea	Efficacy
	-----	-----
CREA	1.00	2.30 (0.67) 3.42
SEFF1	- -	1.00
SEFF2	- -	0.41 (0.07) 6.11
SEFF3	- -	0.63 (0.09) 7.39
SEFF4	0.06 (0.01) 5.49	0.45 (0.09) 4.94

Indirect Effects of ETA on Y

	Crea	Efficacy
	-----	-----
CREA	- -	2.30 (0.67) 3.42
SEFF1	- -	- -
SEFF2	- -	- -
SEFF3	- -	- -
SEFF4	- -	0.14 (0.04) 3.21

Total Effects of KSI on Y

	Person	Socialized
-----	-----	-----
CREA	1.44 (0.12) 11.66	-0.01 (0.17) -0.03
SEFF1	0.21 (0.02) 10.66	-0.02 (0.03) -0.85
SEFF2	0.09 (0.02) 5.43	-0.01 (0.01) -0.84
SEFF3	0.13 (0.02) 6.27	-0.01 (0.02) -0.84
SEFF4	0.16 (0.02) 9.30	-0.01 (0.02) -0.47

注 10：测量模型中，与组织社会化有关的所有整体效应亦未达显著。显示组织社会化变量的效果微弱。

Completely Standardized Indirect Effects of ETA on Y

	Crea	Efficacy
-----	-----	-----
CREA	-	0.33
SEFF1	-	-
SEFF2	-	-
SEFF3	-	-
SEFF4	-	0.14

Standardized Total Effects of KSI on Y

	Person	Socialized
-----	-----	-----
CREA	3.30	-0.01
SEFF1	0.48	-0.03
SEFF2	0.20	-0.01
SEFF3	0.30	-0.02
SEFF4	0.35	-0.01

Completely Standardized Total Effects of KSI on Y

	Person	Socialized
-----	-----	-----
CREA	0.78	0.00
SEFF1	0.75	-0.05
SEFF2	0.32	-0.02
SEFF3	0.39	-0.03
SEFF4	0.57	-0.02

根据所有的效应分析报表,我们可以整理出结构模型当中的所有变量的效应分析,列于表 8.3。值得注意的是,由于整个模型当中同时包含了测量模型与结构模型,因此在整理数据时,应采用完全标准化解的数值,使得所有的参数均具有惯用的标准化性质(数值介于-1 至+1 之间)。

至于测量模型的效应,由于不是研究分析的主要焦点,因此不必详细说明。如果读者想要完成此一部分的报表整理,可以参考先前的范例自行加以整理,在此予以省略。

表 8.3 潜在变量路径分析结构模式各项效果分解说明^a

自 变 量		因变量(内生潜在变量)			
		η_2 自我效能		η_1 创意教学行为	
		标准化系数	t 值	标准化系数	t 值
外 源 变 量	ξ_1 创造性格				
	直接效应	.80	10.66***	.52	5.14**
	间接效应	-	-	.26	3.39***
	整体效应	.80	10.66***	.78	11.66***
	ξ_2 组织社会化				
	直接效应	-.06	-.85	-.02	.31
	间接效应	-	-	.02	-.83
	整体效应	-.06	-.85	.00	-.03
	η_2 自我效能				
内 生 变 量	直接效应			.33	3.42***
	间接效应			-	-
	整体效应			.33	3.42***

注: t 值大于 1.96 时, * $p < .05$; 大于 2.58 时, ** $p < .01$; 大于 3.29 时, *** $p < .001$

^a本表数据系依照修正后的模型所得到的t检验值与完全标准化解编制。

第五节 AMOS 的统合模型分析

本范例所示范的是由 AMOS 读取 SPSS 当中的协方差矩阵数据。如图 8.8 所示。

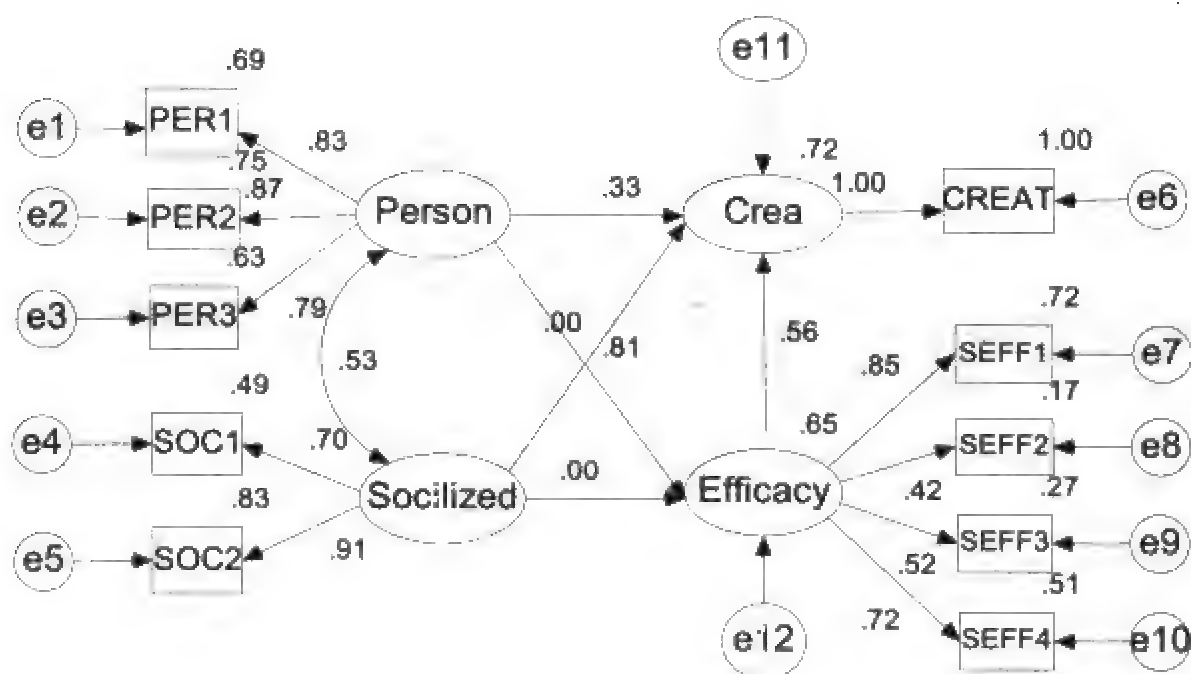
	rowtype_	varname_	CREAT	SEFF1	SEFF2	SEFF3	SEFF4	PER1	PER2	PER3	SOC1	SOC2
1	n		250.00	250.00	250.00	250.00	250.00	250.00	250.00	250.00	250.00	250.00
2	cov	CREAT	17.711									
3	cov	SEFF1	1.843	.404								
4	cov	SEFF2	.801	.146	.374							
5	cov	SEFF3	1.243	.227	.110	.589						
6	cov	SEFF4	1.692	.226	.115	.208	.393					
7	cov	PER1	7.518	1.074	.301	.531	.741	7.565				
8	cov	PER2	7.585	1.062	.538	.609	.806	5.202	7.046			
9	cov	PER3	6.499	.906	.388	.434	.664	4.530	4.626	6.335		
10	cov	SOC1	3.020	.369	.237	.322	.196	1.947	1.524	1.649	4.482	
11	cov	SOC2	2.663	.318	.307	.335	.326	2.092	1.824	1.662	2.335	2.982

图 8.8 SPSS 数据库中的协方差矩阵数据

在 SPSS 的数据库中建立协方差矩阵，仅需将协方差数据直接贴入一个新的 SPSS 窗口中，第一个变量名称输入“rowtype_”，第二个变量名称输入“varname_”，接下来依序输入 10 个变量名称。在第一行打入 n，说明样本数。第一栏打入 cov 以标示为协方差，第二栏打入各变量名称（必须与变量窗口当中的变量名称一致）。然后在下三角区域内贴上或填入协方差数据即可。开启 AMOS 后，完全依照一般程序，即可读入所有的 10 个变量。

一、模型界定

本范例的统合模型依照先前所说明的模式界定，在 AMOS 中直接进行假设图形绘制。值得注意的是，若要以单一指标（crea）反映潜在变量（creat），必须限制误差项（e6）的方差为 0，始可进行模式估计。下图是估计完成后的图示。



二、AMOS 的报表解读

本范例的样本数共计 250 个。变量的摘要显示模型中的所有变量, 包含观察变量(内生变量)与非观察变量(外源变量), 观察变量有 10 个, 非观察变量有 16 个, 共计 26 个变量。模型的纪录显示模型的测量数据数为 55 个, 有 25 个参数被估计, 因此自由度为 30。

本范例 $\chi^2=80.641$, $df=30$, $GFI=.945$, $SRMR=.042$, $RMSEA=.082$, 各拟合指数数据良好, 显示此模型具有理想的拟合度。

从参数估计结果的报表数据可以得知, Person(人格特质)对于 Efficacy(自我效能)与 Crea(创意行为)的结构参数皆达显著, 标准化回归系数值分别为.805 与.334, 表示这两条参数对于 Crea(创意行为)的解释具有统计的意义。此外, Efficacy 对于 Crea 亦有显著的正向影响。因此, Person 能够透过 Efficacy 间接影响 Crea, 其间接效应为.449。从 Squared Multiple Correlations (SMC) 的估计结果, 显示两个内生

变量 (Efficacy 与 Crea) 被外源变量解释的比例分别为.651 与.720。这些数据皆与 LISREL 估计得到的结果相同。

AMOS Output

Model Fit

CMIN
RMR, GFI
Baseline Comparison
Parimony-Adjusted
NCP
FMIN
RMSEA
AIC

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	25	80.641	30	.000	2.688
Saturated model	55	.000	0		
Independence model	10	1257.550	45	.000	27.946

RMR, GFI

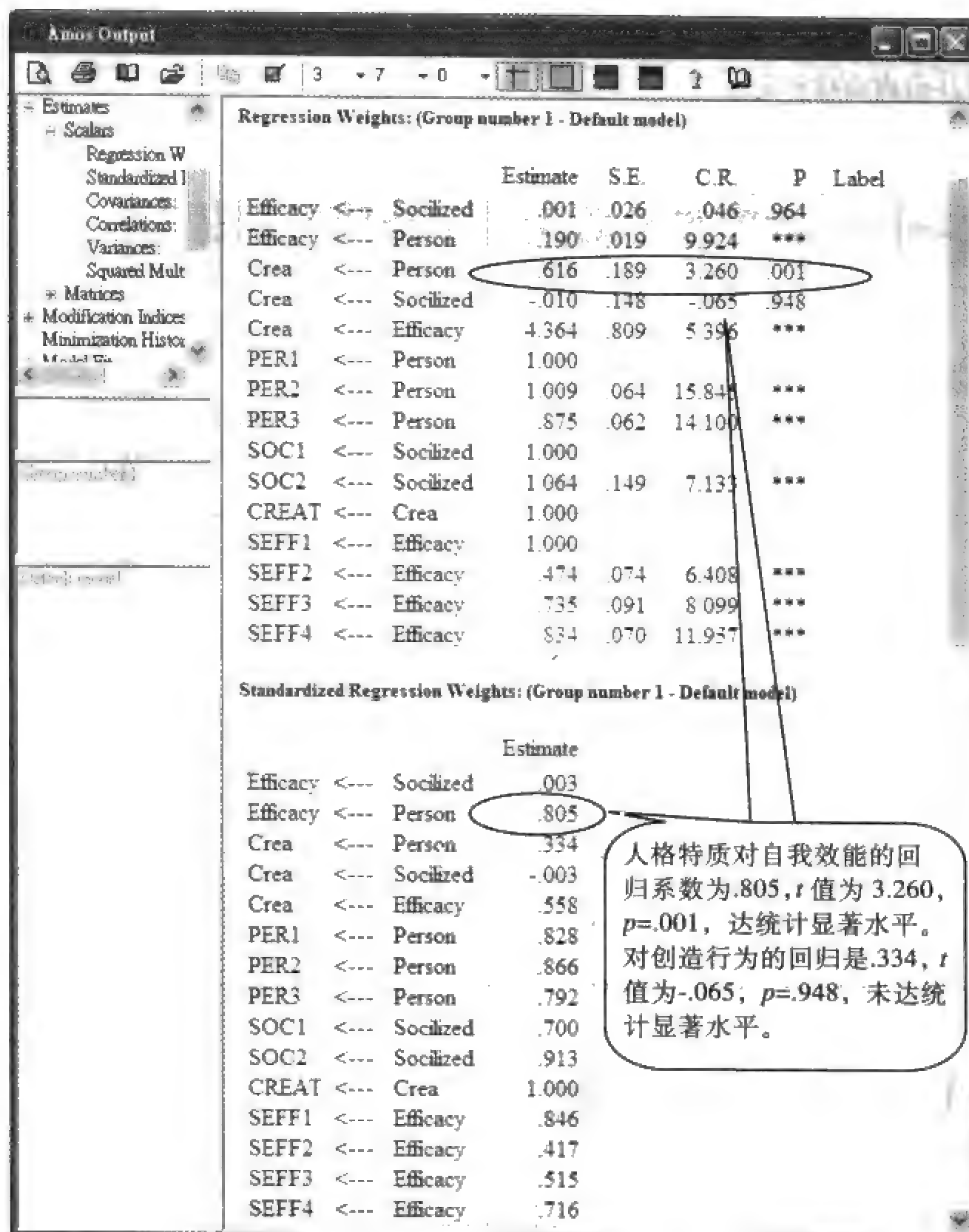
Model	RMR	GFI	AGFI	PGFI
Default model	.104	.945	.899	.515
Saturated model	.000	1.000		
Independence model	2.254	.361	.219	.295

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.936	.904	.959	.937	.958
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Standardized RMR

Default model	Standardized RMR = .0416
---------------	--------------------------



第六节 Mplus 的统合模型分析

一、Mplus语法

```

Title: Ch8 SEM Stage 2 Modified Model using MPLUS5
DATA: File is Ch8.cov;
      Type is covariance;
      NOBSERVATIONS are 250;
Variable:
Names are CREAT SEFF1 SEFF2 SEFF3 SEFF4 PER1 PER2 PER3 SOC1 SOC2;
Model:
  Crea by CREAT;
  Efficacy by SEFF1-SEFF4;
  Person by PER1 PER2 PER3;
  Soclized by SOC1 SOC2;
  Crea on Efficacy Person Soclized;
  Efficacy on Person Soclized;
  CREAT @0;

MODEL INDIRECT:
  Crea IND Efficacy Person;
  Crea IND Efficacy Soclized;
  Crea IND Person;
  Crea IND Soclized;

OUTPUT:
tech1;
STANDARDIZED;

```

间接效应估计

主指令为 Model Indirect:

要求计算机列出间接效应及其标准误。
最终因变量放在 IND 的左边，最初自变量放在同一行的最右边。外源变量仅能放在最右侧。

二、Mplus报表

Ch8 SEM Stage 2 Modified Model using MPLUS5

SUMMARY OF ANALYSIS

Number of groups	1
Number of observations	250
Number of dependent variables	10
Number of independent variables	0
Number of continuous latent variables	4
Observed dependent variables	
Continuous	
CREAT	SEFF1 SEFF2 SEFF3 SEFF4 PER1
PER2	PER3 SOC1 SOC2

```

Continuous latent variables
  CREA          EFFICACY  PERSON      SOCLIZED

Estimator
Information matrix
Maximum number of iterations
Convergence criterion
Maximum number of steepest descent iterations

Input data file(s)
  Ch8.cov

Input data format  FREE

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

          Value          80.964
Degrees of Freedom          30
P-Value          0.0000

Chi-Square Test of Model Fit for the Baseline Model

          Value          1262.601
Degrees of Freedom          45
P-Value          0.0000

CFI/TLI

          CFI          0.958
          TLI          0.937

Loglikelihood

          H0 Value          -3943.495
          H1 Value          -3903.013

Information Criteria

          Number of Free Parameters          25
          Akaike (AIC)          7936.991
          Bayesian (BIC)          8025.027
          Sample-Size Adjusted BIC          7945.775
            (n* = (n + 2) / 24)

RMSEA (Root Mean Square Error Of Approximation)

          Estimate          0.082
          90 Percent C.I.          0.061  0.104
          Probability RMSEA <= .05          0.008

SRMR (Standardized Root Mean Square Residual)

          Value          0.042

```

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
CREA BY CREAT	1.000	0.000	999.000	999.000
EFFICACY BY SEFF1	0.846	0.028	30.736	0.000
SEFF2	0.417	0.057	7.331	0.000
SEFF3	0.515	0.051	10.052	0.000
SEFF4	0.716	0.037	19.533	0.000
PERSON BY PER1	0.828	0.025	33.233	0.000
PER2	0.866	0.022	39.688	0.000
PER3	0.792	0.028	28.123	0.000
SOCLIZED BY SOC1	0.700	0.054	12.922	0.000
SOC2	0.913	0.058	15.827	0.000
CREA ON EFFICACY	0.558	0.093	5.990	0.000
PERSON	0.334	0.100	3.329	0.001
SOCLIZED	-0.003	0.052	-0.065	0.948
EFFICACY ON PERSON	0.805	0.056	14.390	0.000
SOCLIZED	0.003	0.071	0.046	0.964
SOCLIZED WITH PERSON	0.532	0.060	8.855	0.000
Variances PERSON	1.000	0.000	999.000	999.000
SOCLIZED	1.000	0.000	999.000	999.000
Residual Variances				
CREAT	0.000	999.000	999.000	999.000
SEFF1	0.284	0.047	6.087	0.000
SEFF2	0.826	0.047	17.399	0.000
SEFF3	0.734	0.053	13.893	0.000
SEFF4	0.488	0.052	9.306	0.000
PER1	0.314	0.041	7.602	0.000
PER2	0.250	0.038	6.630	0.000
PER3	0.373	0.045	8.366	0.000
SOC1	0.511	0.076	6.739	0.000
SOC2	0.167	0.105	1.581	0.114
CREA	0.280	0.038	7.402	0.000
EFFICACY	0.349	0.060	5.833	0.000

STANDARDIZED TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Effects from PERSON to CREA				
Total	0.783	0.046	17.056	0.000
Total indirect	0.449	0.085	5.303	0.000
Specific indirect				
CREA EFFICACY PERSON	0.449	0.085	5.303	0.000
Direct CREA PERSON	0.334	0.100	3.329	0.001
Effects from SOCLIZED to CREA				
Total	-0.002	0.059	-0.027	0.979
Total indirect	0.002	0.039	0.046	0.964
Specific indirect				
CREA EFFICACY SOCLIZED	0.002	0.039	0.046	0.964
Direct CREA SOCLIZED	-0.003	0.052	-0.065	0.948
Effects from PERSON to CREA				
Sum of indirect	0.449	0.085	5.303	0.000
Specific indirect				
CREA EFFICACY PERSON	0.449	0.085	5.303	0.000
Effects from SOCLIZED to CREA				
Sum of indirect	0.002	0.039	0.046	0.964
Specific indirect				
CREA EFFICACY SOCLIZED	0.002	0.039	0.046	0.964

第七节 结 语

在 SEM 分析架构下，同时包含测量与结构部分的统合模型可以说是最完整的分析形式。这也是 Maccallum 和 Austin (2000) 认为 SEM 最重要的一种应用。换言之，前面章节中所介绍的验证性因素分析与路径分析都只是 SEM 分析的一种特例，两者的结合才是完整的 SEM 应用。至于测量与结构部分是要以一阶段、二阶段或四阶段程序来分别依序处理，只是操作上的考虑。研究者其实可以依照个人研究所需，就不同的分析重点与研究问题进行取舍。

在本章完成了统合模型的介绍之后，有关 SEM 的基本操作程序与相关原理已经告一段落，后面的章节则可以说是统合模型的延伸应用。至于关于 SEM 更进一步的高阶应用与当代最新的发展，基于本书的定位与篇幅限制，则请读者参考当代期刊之中的讨论（例如《Psychological Methods》、《Structural Equation Modeling》）或新书的介绍（例如 Hancock 于 2006 年所出版的《Structural Equation Modeling: A Second Course》）。

曾有学生打趣说，SEM 的学习有三个境界，第一是“知其然”、第二是“知其所以然”、第三是“知其所以不然”。第一个“知其然”阶段的达成，可以从学习者第一次能够顺利得到 SEM 分析结果，而 LISREL 软件终于跳出一个美丽图形作为判断标准。第二个阶段“知其所以然”的境界，则是学习者可以解释各种数据的意义与解释方式，并判断是否存在不合理的数据、是否存在非正定问题或任何警告信息必须处理。本书前述所有章节的介绍，就是希望读者能够顺利成就这两个阶段，以利读者的研究可以顺利完成。

第三个“知其所以不然”的最高境界则是知道隐藏在 SEM 分析背后的盲点、危机或误用的各种可能性，进而能够在适当的情况下正确使用 SEM 这套技术解决相对应的问题，有效提高分析的质量与研究结果的理论实务价值；甚至于可以教导他人学习 SEM、执行 SEM、担任 SEM 的咨询协助工作。要达到这个境界，无法仰赖他人，更无法透过教科书取得系统的获知，而是需要经验的累积，加上一定程度的数理基础作为支撑，与志同道合者的意见交流与相互切磋，持续进行相关议题的研究讨论，才能参透 SEM 背后的玄机与奥妙。以我个人的经验来看，若不坚持个八年十年，要能在 SEM 领域得心应手、左右逢源是不太可能的一件事，真可谓是十年树木、百年树人。虽然代价很高，路途遥远，但若一旦能够达此境界，除个人受益，亦能庇佑他人，更是学术工作应有的坚持实践。

INDEX

[illegible][illegible]

第九章

多样本结构方程模型

前面几章有关结构方程模型的应用,都仅限于单一一个样本的共变结构所进行的分析。也就是说,SEM模型所分析的数据系来自于总体其中的一个随机样本的观察数据。但是,在很多情形下,研究者必须处理来自多个样本的数据,例如不同类型的工作者(内勤人员与外勤人员)、不同层级的员工(主管与部属)、不同地区人群的比较(台湾样本与大陆样本)、性别差异(男性与女性)、不同实验状况(实验组与对照组)。此时,为了检验研究者所提出的假设模型在不同样本间是否相等(invariant or equivalent),也就是检验多样本恒等性或等同性(multi-group invariance)时,单一样本的SEM分析即不敷所需,而必须使用多样本结构方程模型(multi-group structural equation modeling)进行共变结构的分析。

第一节 多样本分析的概念

一、多样本分析的基本概念

简单来说,多样本结构方程模型与单样本分析的主要差异,在于原有的SEM分析之外,增加了另一个(或多个)平行样本的估计。换句话说,就是在SEM模型的基本设定之外,另外纳入一个类别变量反映共变结构的跨样本变化。也即由于模型需同时考虑两个或两个以上的样本共变结构,研究者必须提供这些有待检测的各个样本的观察数据,以SEM模型来检验两个共变结构是否相同或相异。

举例来说,当我们发展一套组织气氛量表,利用一批样本进行CFA分析之后,我们可能可以继续讨论,是否这套量表在测量不同产业时,有一样的意义?此时即必须纳入一个产业类别变量,将原来组织创新气氛量表的共变结构区分成不同的产业来比较。

或者,当老师关心自我效能、社会期望两者是否会影响学生的成就动机,进而影响学生的学业表现时,这个路径模型就必须纳入一个类别变量“性别”,检验两个样本的结构模型在男生样本与女生样本上是否一样或有所差异。当如果考虑的是学生的不同年级,那么即成为多个样本(一、二、三年级样本)的SEM模型比较。

虽然多样本 SEM 仅仅是在原有的 SEM 模型中增加了一个类别变量,但是由于 SEM 模型本身有多种组合(例如测量模式、结构模式、混合模型)以及各种不同的参数(路径参数、因素载荷参数、相关系数等),因此 SEM 模型的多样本恒等性评估,也就存在着多种不同的变化。Byrne (1994)指出多样本之间的恒等性检验至少有下列五种不同的评估途径:

1. 观察变量的数据在不同的样本间是否等同?此一问题所涉及的问题是测量模式的恒等性评估。
2. 研究者所提出的因素结构在不同的样本间是否等同?此一问题所涉及的问题是结构模式的恒等性评估。
3. 在结构方程模型中,是否有某些特定的参数(如回归系数或相关系数)在不同样本间等同?此一问题所涉及的问题不是整个模型的恒等性问题,而是比较特定参数在不同样本的估计状况。
4. 潜在变量的平均数在不同的样本间是否等同?此一问题涉及了结构方程模型的平均数结构评估。
5. 因素结构在不同样本上是否可以复制?也就是复核效化(cross-validation)的概念。

结构方程模型延伸到多样本间的应用,虽然是 SEM 技术发展 to 相当程度后的一种应用,但并不是一种新的概念,早在 20 世纪 40 年代即有学者提出测量恒等性(measurement invariance)的概念(Thurstone, 1947)。在多样本分析的架构下,测量恒等性就是测量模型在不同样本间是否相同的一种应用。另一方面,结构模型在不同样本间等同与否,则是一种调节效应(moderation)的概念。以 SEM 的术语来说,就是在不同条件下,各结构参数是否相等或不等的检测。也就是说,多样本结构方程模型分析不但可以在测量模型上进行应用,检验跨样本的测量信效度(也就是一种复核效化的应用),更可以延伸到结构参数的跨样本检验,进行调节效应的分析。这更扩大了 SEM 应用的范畴。

有关多样本分析的具体技术直到近一二十年才有快速发展。一方面是因为 SEM 本身学理与概念的成熟,另一方面则是因为计算机硬件与分析软件的推陈出新,使得社会科学研究者得以对于复杂的因素建构与其参数估计进行实证的统计操作。

二、多样本分析的应用：测量恒等性

多样本 SEM 分析可以说是用来评估测量恒等性最佳策略。所谓测量恒等性是指测验工具或评量方法施于不同的对象或于不同的时点上使用时，测验分数或评量结果应具有一定的恒等性 (Reise, Widaman, & Pugh, 1993)，也就是说，当研究者利用一组测验题目测得一个心理的概念（例如焦虑或自尊）并应用于组间的比较（如男女性别或不同年级）时，研究者必须假设测验分数背后的项目分数与尺度对不同的受试对象具有相同的意义 (Drasgow & Kanfer, 1985)。

测量恒等性最具体的应用，是古典测量理论中，对于测验信度的检验，也就是指测验项目的稳定性与可靠性。一般而言，测验编制者希望测验分数不会因时点的变动或施测对象的不同而变化，此即测验形式等同性 (formal identity) 或项目恒等性 (item equivalence) (Hui & Triandis, 1985)。

在 SEM 发展之前，传统的信度估计是以测验得分的波动状况或相关情形为基础，或求取项目间的相关（如内部一致性系数）。多样本 SEM 发展后，测量恒等性除了可以具体应用于信度估计程序探讨测验项目的等同性外，并可以将恒等性的概念延伸到项目的尺度、文字使用的讨论，探讨功能与概念上的恒等性 (functional/conceptual equivalence)，也就是测验项目与其所对应的潜在因素之间的共变结构上 (covariance structure) 的相似性与差异性。例如，同一个量表被转译为不同的语言之后，项目的等同性亦为测量恒等性的一种研究典型范例，称之为转译恒等性 (translational equivalence, or translation equivalence) (Brislin, 1986; Hogevar & El-Zahhar, 1992)。对于一个测验的因素结构在不同研究对象中的相似性则被称为因素恒等性 (factorial invariance)。在跨文化研究中，测量恒等性的检测更是重要，因测量工具应用在跨文化议题时，亟须一套客观的验证程序确保测量的可比较性 (Byrne & Campbell, 1999; Cheung, 2008; Vandenberg & Lance, 2000)。

传统上检验测量恒等性的方法是应用探索性因素分析，例如分别对两个样本执行两次因素分析，比较各自的因素结构（例如，Ahmavaara, 1954; Cliff, 1966; Meredith,

1964; Mulaik, 1972; Please, 1973; Taylor, 1967) 或求取两次因素分析得到的负荷值的相关系数 (Cattell & Baggaley, 1960; Kaiser, Hunka, & Bianchini, 1971; Reynold & Harding, 1983)。这些检验方式受限于 EFA 的限制, 因素结构缺乏明确的判定标准, 无法事先指定因素结构的形态, 而因素结构的相同或相异, 仅能以目测比对或藉建立其他指标的方式进行, 缺乏标准化数据的客观检验。因此, 测量恒等性的评估容易受到技术本身不稳定的影响, 操作上相当困难。再者, 由于一个因素结构系由不同的部分 (factor loadings, factor variances, error variances) 组成, 此种方法无法分就各个不同的恒等性部分加以检测 (Marsh & Hocevar, 1985)。

三、多样本分析的应用：复核效化

如果说测量恒等性所重视的是测量的题目质量与因素结构内在信效度的跨样本恒等性, 那么复核效化即是测量的外在效度的跨样本恒等性。所谓复核效化 (cross-validation) 是指假设模型所具有的跨样本或跨情境的有效性。这个概念很早就受到重视, 早期是应用于回归分析的一种统计技术, 用于检测回归模型的稳定性 (Mosier, 1951; Cattin, 1980), 但随着测量理论与统计技术的发展, 凡是测量分数在不同情境下进行预测与解释的稳定性的检验, 都可以视为复核效化的一种做法。

在结构方程模型中, 学者多将复核效化定义成 SEM 模型 (包括测量模式与结构模式) 在不同样本可重复估计的稳定程度 (Diamantopoulos & Siguaw, 2000), 因此, 复核效化的定义涵盖了测量恒等性的概念。复核效化重视整体模式的跨样本效用, 测量恒等性则强调测量的个别参数与因素结构的稳定与否。从某一个研究当中所证实的最佳模型, 在该研究样本上虽有最佳的拟合度, 但并不代表在其他样本也具有相当程度的拟合度。这里所谓的其他样本, 可能是与该研究相同总体的另一个样本, 也可能是不同总体下的另一个独立样本。

复核效化的典型做法是利用多样本模型分析针对同一个假设模型在两个样本上进行估计, 其中一个样本称为测定样本 (calibration sample), 另一个样本则为效度样本 (validation sample) (Kaplan, 2000)。首先将被检验的 SEM 模型在测定样本上进行估计后, 将所估得的参数套用在效度样本上, 然后进行模型拟合检验。

如果研究者先后两次获得同一个总体的两个随机样本，此时即可利用这两个样本进行同一个模型是否等价的模型稳定性（model stability）检验。如果研究者无法重复取样两次，那么一种变通的做法是将整个研究的样本随机分割成两个子样本以进行复核效化的估计程序。一般来说，SEM 研究所使用的样本数少则数百，多则数千，因此研究者多半拥有足够规模的样本来切割样本。

当研究者的两个样本系来自不同的总体时，等于是将假设模型更进一步扩展到其他不同总体下，此时复核效化所检验的不仅是模型的稳定性，而是效度延展性（validity extension）的检验。

进一步的，如果某一个研究中，研究者有多个模型同时进行竞争比较，若再利用同一个总体的另一个样本进行重复检验，此时是 SEM 多重样本分析的模型选择性（model selection）检验，其优点是模型的选择程序具样本间的稳定性。如果所使用的是不同总体的样本，此时即可称为测量效度在不同情境下的效度类化（validity generalization）检验。Diamantopoulos 与 Siguaw（2000）将这四种不同的复核效化程序加以整理，得到一个完整的概念，如表 9.1 所示。

表 9.1 复核效化的不同形式与检验内容

研究中的模型数目	效度样本来源	
	相同总体	不同总体
单一模型	模型稳定性 (Model stability)	效度延展性 (Validity extension)
多重模型	模型选择性 (Model selection)	效度类化性 (Validity generalization)

修改自 Diamantopoulos 与 Siguaw（2000），p.130。

第二节 多样本分析的统计原理

由于 SEM 模型牵涉到因素结构的设定,因此多样本 SEM 首要问题是因素结构的设定在样本间的可比较性。较普遍的做法是 Jöreskog (1971) 所提出渐近模型的比较程序评估多样本因素结构恒等性。也就是将跨样本因素结构模型视为单一样本的因素结构的更严格限制模型,再透过嵌套模型的比较决定样本间的因素恒等性。

一、多样本结构方程式

首先,对于多样本的测量模型因素结构,我们可以用下列的通式表示:

$$x_g = \Lambda_{xg}\xi_g + \delta_g \quad (9.1)$$

$$y_g = \Lambda_{yg}\eta_g + \varepsilon_g \quad (9.2)$$

其中 x_g 与 y_g 是第 g 个样本的观察变量向量, Λ_{xg} 与 Λ_{yg} 是第 g 个样本对应于观察变量的因素载荷矩阵, ξ_g 与 η_g 是潜在变量向量, δ_g 与 ε_g 是测量残差向量。结构模型的通式如下:

$$\eta_g = B_g\eta_g + \Gamma_g\xi_g + \zeta_g \quad (9.3)$$

二、多样本模型分析的估计原理

(一) 多样本拟合函数

基于上述方程式,多样本 SEM 分析的参数估计是透过下列拟合函数的最小化来获得的:

$$F = \sum_{g=1}^G \frac{N_g}{N} F_g(S_g, \Sigma_g, W_g) \quad (9.4)$$

其中 F_g 为第四章所介绍的各种不同估计程序的拟合函数, N_g 为各组下的样本数,显示 F 函数考虑到不同样本之间的样本数的影响。整个模型的卡方值为总样本数乘以估计拟合函数值,而整个模型的自由度为:

$$df = G[\frac{1}{2}(p+q)(p+q+1)] - t \quad (9.5)$$

整个模型的自由度为总测量数据点数 (DP) 减去各样本估计参数的总和 (t)。所以在多样本 SEM 分析当中,只会产生一个卡方估计数与一个自由度,表示多样本分析的整体拟合情形,而 GFI、RMR 等指数则是分就不同的样本报告之。

(二) 最大概似法

ML 法是 SEM 分析最常用的方法,因此,本节以 ML 法为基础,说明多样本 SEM 分析的原理。

当多个样本之间为独立样本时,且各样本来自于正态总体,此时第 g 个样本的对数概率函数 (log-likelihood function) 可以写为:

$$\log L_0(\Omega)_g = -\frac{n_g}{2} \log |\Sigma_g| + tr(S_g \Sigma_g) \quad (9.6)$$

将各样本的对数概率函数相加总，得到下式：

$$\log L_0(\Omega) = \sum_{g=1}^G \log L_0(\Omega) \quad (9.7)$$

取上二式的最小化，即得到最大概似法的拟合函数 F_{ML} (maximum likelihood fitting function)：

$$F_{ML} = \log|\Sigma| + tr(S\Sigma^{-1}) - \log|S| - q \quad (9.8)$$

三、多样本参数估计

多样本 SEM 分析的参数估计，在参数的标准化处理上与单样本 SEM 分析的做法有所不同。当研究者所关心的跨样本比较时，变量的标准化必须是跨样本来进行，也就是建立跨样本的共同计量尺度 (common metric for all groups)，而不是单独针对某一个样本下的潜在变量或观察变量标准化。

例如，当不同样本下，测量变量与潜在变量的方差不相等时，因素载荷参数 (Λ_x 、 Λ_y) 若在各自的样本下被标准化，跨样本间的参数比较即缺乏一个共同计量尺度，被设定为恒定的因素载荷的估计数的标准化是不完全的。

一般在单样本 SEM 分析中，会得到三种参数解：非标准化估计值、标准化估计值 (standardized solution) (将潜在变量进行标准化) 与完全标准化估计值 (completely standardized solution) (将潜在变量与观察变量同时进行标准化)。由于仅有一个样本，因此这三种解是以所有样本来计算的。但是在多样本分析中，则有全体样本、个别样本两种分析条件，因此参数估计的标准化程序更加复杂。在 LISREL 报表中，会出现 SOLUTION STANDARDIZED TO A COMMON METRIC 标准化解，表示参数估计解系将潜在变量的测量尺度经过跨样本标准化，亦即具有跨样本的共同计量尺度。相对的，个别样本下的参数估计标准化解，称为组内标准化解 (WITHIN GROUP STANDARDIZED SOLUTION)，则会与跨样本标准化解不同。

在多样本分析中, 完全标准化解也区分为跨样本的共同量尺完全标准化解 (COMMON METRIC COMPIETELY SOLUTION STANDARDIZED) 与组内完全标准化解 (WITHIN GROUP COMPLETELY STANDARDIZED SOLUTION)。这两种标准化解是将测量变量与潜在变量的方差都设为 1, 所得到的参数均具有标准化的特性。

四、恒等性检测策略

(一) 协方差矩阵恒等性检验

由于 SEM 的分析是以协方差矩阵为基本材料, 因此多样本间的模型检验, 第一个需要评估的是样本间的协方差矩阵的等同性是否成立。此时, 尚未有任何的假设模型被设定在 SEM 模型中。所检验的假设列举如下:

$$H_{\Sigma} : \Sigma_1 = \Sigma_2 = \dots = \Sigma_G$$

利用 Box's M 检验 (Tatsuoka, 1988), 可以检验上述假设, 检验的统计量计算式如下:

$$M = n \log |S| - \sum_{g=1}^G n_g \log |S_g| \quad (9.9)$$

由于 M 统计量的性质与卡方分布相似, 因此 Box's M 检验可以利用卡方分布进行显著性检验, 自由度如下:

$$df_{\Sigma} = \frac{1}{2}(g-1)q(q+1) \quad (9.10)$$

当显著性检验未达显著水平, 显示 H_{Σ} 假设没有被推翻, 多样本间的协方差矩阵完全相同, 此时应停止多样本的 SEM 分析, 将各样本的数据加以整合, 改以单一样本 SEM 分析。如果显著性检验达显著水平, H_{Σ} 假设被推翻, 表示样本间的协方差矩

阵具有差异,此时可继续进行不同类型的多样本 SEM 分析,来探讨样本间的差异究竟是发生在何处。在此,Box's M 检验的性质就如同整体检验 (overall test)。

(二) 因素恒等性检验

当 Box's M 检验指出样本间的共变结构具有差异后,即可以开始检验不同样本的差异。首先应检验的是测量模式的差异,包括两个步骤,第一是因素的结构是否相等,也就是检测因素的数目 (k) 是否相等,假设为:

$$H_k: k_1 = k_2 = \dots = k_G$$

此一假设的考验可以视为是 G 次独立的未限制因素个数的因素分析模型检验,每一个样本进行一次卡方检验,个别检验的自由度为:

$$df_k = \frac{1}{2}[(q-k)^2 - (q+k)] \quad (9.11)$$

G 次卡方检验的总自由度,将上式的 df_k 乘以 g 即可以得到。如果此一假设遭到拒绝,显示因素个数不相等,跨样本的检测应该终止,改以个别的样本进行单一样本 SEM 估计,以了解个别样本下的因素结构应如何。

当 H_k 假设被接受后,可以进行第二个步骤,考验因素载荷是否相等,也就是因素载荷的数值是否相等。假设如下:

$$H_A: \Lambda_1 = \Lambda_2 = \dots = \Lambda_G$$

操作上,此一假设的检验是利用一组嵌套模型进行卡方差异检验,限定模型(恒等模型)是将各样本的 SEM 测量模型中的 Λ 矩阵设定为样本间相等,其他的参数则不做任何设定,基准模型(baseline model)则是 Λ 矩阵与其他所有参数均未设定样本间相等。将两个模型分别进行参数估计后,相减得到卡方值差异数,达显著表示恒等模型能够反映数据,可知恒等性存在,可以继续进行其他各参数(测量残差、因素间相关系数、残差相关系数)的恒等性检验,例如同时考验因素载荷与残差变异跨样本恒等性 ($H_{\Lambda\Theta}$),以及同时考验因素间相关、因素载荷与残差变异的跨样本恒

等性 ($H_{\Lambda\Theta\Phi}$)。任何一个假设的卡方值差异数未达显著, 显示参数的恒等性不存在, 研究者应停止 SEM 分析, 讨论所检验的参数的样本间差异发生在何处。

Jöreskog (1971) 再提出上述检验程序时, 指出了多次假设检验可能带来的第一类错误几率膨胀等潜在问题, 因此建议实际操作时可以不严格遵循这些渐近的检验步骤, 直接针对研究者所关心的议题加以检验即可。一般来说, 对于因素个数与因素载荷恒定性的检验可以整合为一, 称为因素恒等性 (factorial invariance) 检验。大多数研究对于测量恒等性的检验, 都是针对因素载荷来检验, 其他的各参数恒等性则视研究者的需要检测之。

(三) 其他参数恒等性检验

因素恒等性的检验主要是针对测量模型内的各参数进行样本间的等同性检验。藉由相同的原理, 当因素恒等性假设成立的情况下, 研究者可以进一步检验结构模式当中各参数的恒等性, 例如 Beta、Gamma、Psi 矩阵各参数的恒等性。

值得注意的是, 当结构模型当中具有潜在变量的设定时 (例如潜在变量路径分析, 或是高阶验证性因素分析), 结构参数的恒等性必须在因素个数与因素载荷恒等性存在的前提下才能进行, 否则结构参数的恒等性即缺乏了测量恒等性作为基础的合理性。相对的, 当结构模型当中没有任何潜在变量的设定时, 例如观察变量的路径分析, 研究者可以直接进行结构参数的样本恒等性检验。

除了针对整个矩阵进行恒等性的考验, 研究者可以针对矩阵中的特定部分进行检验, 称为部分恒等性检验 (test for partial invariance) (Byrne, Shavelson, & Muthén, 1989)。其步骤是当被检验的矩阵在不同样本间具有差异的情况下, 再考验究竟矩阵中的何者或哪些参数具有跨样本的差异性, 此种考验模式称为部分恒等性检验。

第三节 多样本分析：测量恒等性检验

在第五章的范例中，我们说明了“组织创新气氛量表”十八题短题本的验证性因素分析的具体做法，该量表具有“组织价值”、“工作方式”、“团队合作”、“领导风格”、“学习成长”、“环境气氛”六个因素，每个因素（分量表）由三个题目组成。在本范例中，我们则纳入了性别变量，将 CFA 分析扩展到性别之间的恒等性比较，用以证明组织创新气氛的知觉的因素结构在不同性别之间是否具有恒等性。

以跨样本 SEM 模型分析进行因素恒等性检验，必须使用一系列的模型比较程序，区分为两个阶段逐步检验不同层次的恒等性假设。第一个阶段为单样本 CFA 检验程序，第二个阶段为恒等性检验，也就是应用跨样本 SEM 模型分析进行的检验程序。

第一个阶段首先由以全体样本估计所得的良好拟合 CFA 模型为基础，进一步分就不同样本，进行个别样本下的 CFA 分析。如果有两个样本，必须进行两次独立的 CFA 分析，如果有三个样本，就必须进行三次独立的 CFA 分析。这些独立的 CFA 分析，也必须具有一定的拟合度。如果有哪一个样本的 CFA 检验显示该 CFA 模型在该总体下拟合度不理想，研究者应停止跨样本的比较，直接进行各自样本下的 CFA 检验。此时，因素恒等性是不存在的。

第二个阶段是当研究者藉由前一个阶段的检验，确认出一个整体的 CFA 拟合模型后，再利用跨样本程序探讨该 CFA 模型的跨总体恒等性，据以证明因素之间具有测量的恒等特性。具体做法是建立一个阶段性逐步设限的嵌套模型来一一检验各种恒定假设下的模型拟合度的变化。以下，我们仅将第二阶段的跨样本 SEM 模型分析的做法加以详述，有关第一个阶段的单样本 CFA 分析，则不列出详细操作流程，读者可以参考前面章节的范例说明。

一、假设模型的建立

跨样本 SEM 分析用于因素恒等性的检验,是由下列的各假设模型的分析与比较来进行的:

模型一:基准模型,跨样本之间没有任何恒等假设,因素结构则假设相等。基准模型是两个独立无关联但结构相同的 CFA 模型之组合,其卡方值为两个个别样本以同一个因素结构进行估计的总和。

模型二:因素载荷恒等模型,因素结构在不同总体间被假设为相同,同时因素载荷具有跨样本的恒等设限。也就是两个样本的 Λ_x 矩阵假设为相等同。

模型三:因素载荷与测量残差变异恒等模型。此模型较模型二更进一步假设测量残差变异量具有跨总体的恒等性。也就是两个样本的 Λ_x 矩阵与 Θ_ϵ 矩阵被假设为相等同。

模型四:因素载荷、测量残差变异、因素方差与协方差均恒等模型。此模型较模型三更进一步假设因素的方差与共变量具有跨总体的恒等性。也就是两个样本的 Λ_x 矩阵、 Θ_ϵ 矩阵与 Φ 矩阵被假设为相等同。此时,由于所有的参数都被设定为相等同,因此可以视为完全等同模型。

利用这四个模型,我们可以进行不同层次的因素恒等性检验,当越多的参数被设定为恒等,也就反映了因素恒等性越强。这四个模型由于被估计的参数越来越少,因此自由度越来越大,卡方值也会越来越大。利用卡方差异检验,可以决定新增的设限是否具有显著的影响,达显著的卡方差异检验,代表所检验的恒等性是存在的。相对的,卡方差异检验未显著时,代表所检验的恒等性不存在。

二、模型界定

在范例五当中,我们已经指出 18 个测量变量可以产生 171 个测量点数($DP=171$),18 个测量变量测量六个因素的知觉强度的 CFA 模型,共有 51 个被估计的参数,设定情形如下:

1. 模型中有 18 个测量变量(V1 至 V18,为内生观察变量,亦作因变量)与 6 个潜在变量(F1 至 F6,为外源潜在变量,亦作自变量)。

2. 模型中有 18 个测量残差，其变异量被自由估计。
3. 为了使 6 个潜在变量（因素）的量尺得以确立，因此每一个因素的方差被设定为 1.00。
4. 因素之间的共变则允许自由估计。
5. 测量残差之间被视为独立而没有共变关系。
6. 每一个测量变量仅受单一潜在变量影响。

上述设定条件在第一阶段单样本分析中，对于全体样本、个别样本的个别 CFA 分析都是相同的。

在第二阶段当中，由于同时包括两个样本的两套 CFA 模型进行分析，因此总 DP 值为 $171 \times 2 = 342$ ，而四个被检验模型的设定条件，则有下列之变动：

模型一：两个样本进行 CFA 分析，但是模型中未设定恒等性，模型界定条件未改变。CFA 的 DP 值为个别 CFA 的总和（ $171 \times 2 = 342$ ），总估计参数为 $51 \times 2 = 102$ 。自由度为 240。（Ch9amg1.ls8）

模型二：新增因素载荷恒等性假设后，样本间的因素载荷参数的每一组配对均设定为相等，因此因素载荷参数估计数目为整合模式的 36 个参数的一半（18 个），表示释放了 18 个自由参数，总估计参数较前一个模型减少 18，为 84。自由度为 258。（Ch9amg2.ls8）

模型三：新增残差方差恒等性假设，样本间的残差变异量的每一组配对均设定为相等，因此残差变异量参数估计数目为整合模式的 36 个参数的一半（18 个），表示再释放了 18 个自由参数，总估计参数较前一个模型减少 18，为 66。自由度为 276。（Ch9amg 3.ls8）

模型四：新增因素方差与协方差恒等性假设，样本间的因素方差与协方差参数也少了一半，六个因素的恒等性设定，共减少了 15 个方差协方差估计参数，再释放 15 个自由参数后，总估计参数为 51（等于个别模型的参数估计数）。自由度为 291。（Ch9amg 4.ls8）

三、参数估计

(一) 多样本模型一：未限定 CFA 语法（基准模型）

第一阶段所需进行的单样本 SEM 分析，其 LISREL 语法与第五章范例完全相同，读者可以自行参阅。个别样本进行独立的 CFA 分析时，语法仅有样本数的数字不同，其余皆相同。

表 9.2 列出了模型一进行跨样本 CFA 分析的 LISREL 与 SIMPLIS 语法。语法皆由两个独立的部分所组成，其中 LISREL 语法的第 1 至 11 行用以设定男性样本 CFA 模型分析条件，第 13 至 22 行用以设定女性样本 CFA 模型分析条件，两者除了样本数与数据文件的文件名不同以外，其余指令完全一样。显示两个样本的因素结构是完全相同的设定。

表 9.2 多样本验证性因素分析语法

LISREL 语法范例(Ch9amg1.ls8)	
1	GROUP: MALE
2	DATA NG=2 NI=18 NO=172
3	KM SY FI=Ch9am.cor; ME FI=Ch9am.cor; SD FI=Ch9am.cor
4	LA; A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3
5	MODEL NX=18 NK=6 PH=SY,FR TD=DI,FR LX=FU,FI
6	LK; VALUE JOBSTYLE TEAMWORK LEADERSHIP LEARNING ENVIRONMENT
7	FREE LX(1,1) LX(2,1) LX(3,1) LX(4,2) LX(5,2) LX(6,2) LX(7,3) LX(8,3) LX(9,3)
8	FREE LX(10,4) LX(11,4) LX(12,4) LX(13,5) LX(14,5) LX(15,5) LX(16,6)LX(17,6) LX(18,6)
9	VALUE 1 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
10	PD
11	OUTPUT
12	GROUP: FEMALE
13	DATA NO=140
14	KM SY FI=Ch9af.cor; ME FI=Ch9af.cor; SD FI=Ch9af.cor
15	LA; A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3
16	MODEL NX=18 NK=6 PH=SY,FR TD=DI,FR LX=FU,FI
17	LK; VALUE JOBSTYLE TEAMWORK LEADERSHIP LEARNING ENVIRONMENT
18	FREE LX(1,1) LX(2,1) LX(3,1) LX(4,2) LX(5,2) LX(6,2) LX(7,3) LX(8,3) LX(9,3)

```

19  FREE  LX(10,4) LX(11,4) LX(12,4) LX(13,5) LX(14,5) LX(15,5)
    LX(16,6) LX(17,6) LX(18,6)
20  VALUE 1 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
21  PD
22  OUTPUT
23  End of Problem

```

SIIMPLIS 语法范例(Ch9amg1.spl)

```

1  GROUP: MALE
2  Observed variables: A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3
3  correlation matrix from file Ch9am.cor
4  Sample Size = 172
5  Latent Variables  VALUE JOBSTYLE TEAMWORK LEADERSH LEARNING
   ENVIRONM
6  Relationships
7  A1-A3= VALUE
8  B1-B3 = JOBSTYLE
9  C1-C3 = TEAMWORK
10 D1-D3 = LEADERSH
11 E1-E3 = LEARNING
12 F1 -F3= ENVIRONM
13 Set the Variance of VALUE-ENVIRONM to 1.00
14 Path Diagram
15 GROUP: FEMALE
16 Observed variables: A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3
17 correlation matrix from file Ch9af.cor
18 Sample Size = 140
19 Latent Variables  VALUE JOBSTYLE TEAMWORK LEADERSH LEARNING
   ENVIRONM
20 Relationships
21 A1-A3= VALUE
22 B1-B3 = JOBSTYLE
23 C1-C3 = TEAMWORK
24 D1-D3 = LEADERSH
25 E1-E3 = LEARNING
26 F1 -F3= ENVIRONM
27 Set the Variance of VALUE-ENVIRONM to 1.00
28 Set the Covariance of JOBSTYLE and VALUE Free
29 Set the Covariance of TEAMWORK and VALUE Free
30 Set the Covariance of TEAMWORK and JOBSTYLE Free
31 Set the Covariance of LEADERSH and VALUE Free
32 Set the Covariance of LEADERSH and JOBSTYLE Free
33 Set the Covariance of LEADERSH and TEAMWORK Free
34 Set the Covariance of LEARNING and VALUE Free
35 Set the Covariance of LEARNING and JOBSTYLE Free

```

36	Set the Covariance of LEARNING and TEAMWORK Free
37	Set the Covariance of LEARNING and LEADERSH Free
38	Set the Covariance of ENVIRONM and VALUE Free
39	Set the Covariance of ENVIRONM and JOBSTYLE Free
40	Set the Covariance of ENVIRONM and TEAMWORK Free
41	Set the Covariance of ENVIRONM and LEADERSH Free
42	Set the Covariance of ENVIRONM and LEARNING Free
43	Set the Error Variance of A1-F3 Free
44	End of Problem

对于多样本的设定，LISREL 语法藉由第 2 行中的 NG=2 指令来辨认 LISREL 程序当中具有两个独立的 LISREL 语法程序，然后每一个样本由 GROUP 指令作为开端，例如第 1 行的 GROUP: Male 与第 12 行的 GROUP: Female 用来说明不同样本的语法程序。

表 9.2 当中下半段的 SIMPLIS 语法的第 1 至第 14 行用以设定男性样本 CFA 模型，第 15 至第 43 行设定女性样本 CFA 模型。比较特别的是，由于第一个模型对两个样本并没有设定相同，因此 SIMPLIS 语法必须在第二个样本上，详细列出潜在变量与残差的方差与协方差的设定（由第 27 到第 43 行），如果没有设定，那么所有的方差与协方差都会被设定为等值。

（二）恒等性参数限定语法

对于参数的恒等限定，LISREL 语法的做法是在第一个样本中，以 FR 指令设定该参数为自由估计，然后在第二个或所有的后续样本中以 EQ 指令进行恒等限定。另一种做法是在 Model 指定中，在所需设定的矩阵（例如 LX、LY、PS 等）之后加入下列副指令：

- 1. SP (same pattern)：表示样本间矩阵具有相同的形态。
- 2. SS (same starting value)：表示参数估计使用与前一个样本相同的起始值来进行估计。
- 3. PS (same pattern and starting value)：表示样本间矩阵具有相同的形态并使用与前一个样本相同的起始值来进行估计。

4. IN (invariant): 表示矩阵为恒等矩阵, 矩阵的形态与参数的估计状态 (FR 或 FI), 以及参数数值均与第一个样本相同。

以本范例的模型二为例, 由于增加了因素载荷恒等假设的设定, 必需在第二个样本的模型界定指令 (Model) 当中, 增加一个 $LX=IN$ (Λ_x 矩阵相等) 的指令, 然后移除 FR 与 FI 指令中对于 LX 参数的设定即可。也就是表 9.2 的指令中, 第 18 和第 19 行两行指令删除, 改以在 MO 指令中增加 $LX=IN$ 取代之, 语法变动部分如下:

```
16 MO NX=18 NK=6 PH=SY,FR TD=DI,FR LX=IN
17 LK; VALUE JOBSTYLE TEAMWORK LEADERSHIP LEARNING ENVIRONMENT
20 VALUE 1 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
21 PD
22 OUTPUT
23 End of Problem
```

如果是 SIMPLIS 语法, 只要把第 21 到第 26 行移除, 即可被识别为第二个样本的因素载荷与第一个样本等值。

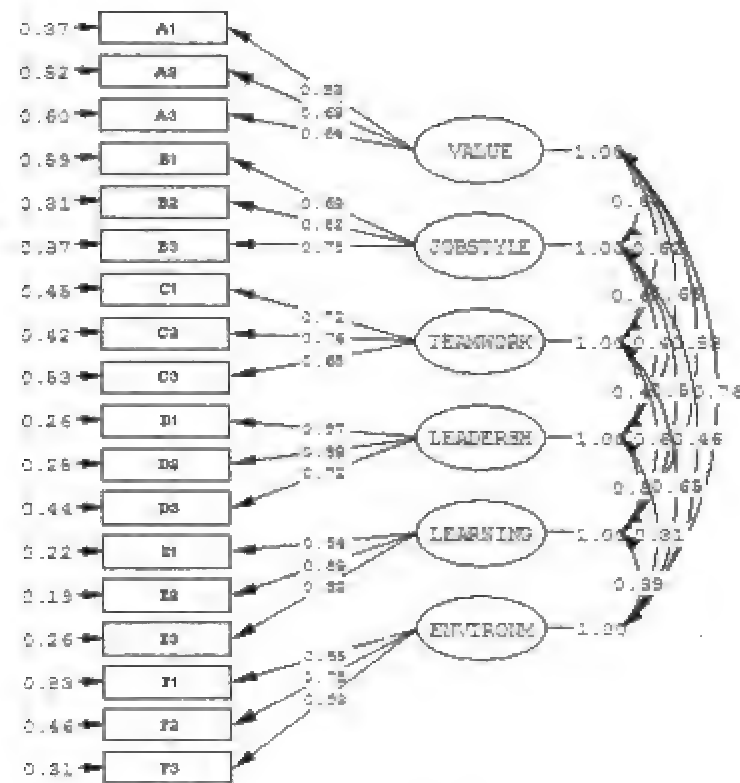
以 LISREL 与 SIMPLIS 语法估计得出的男女生样本参数估计结果如图 9.1, 由图中数据可知, 两个样本在因素载荷上数据相同, 但是测量残差则不同。

模型三的设置依此类推。于 $LX=IN$ 之外, 另外增加一项 $TD=IN$ 设定, 也就是表 9.2 指令的第 19、第 20、第 21、第 22 行指令改以下列指令取代之:

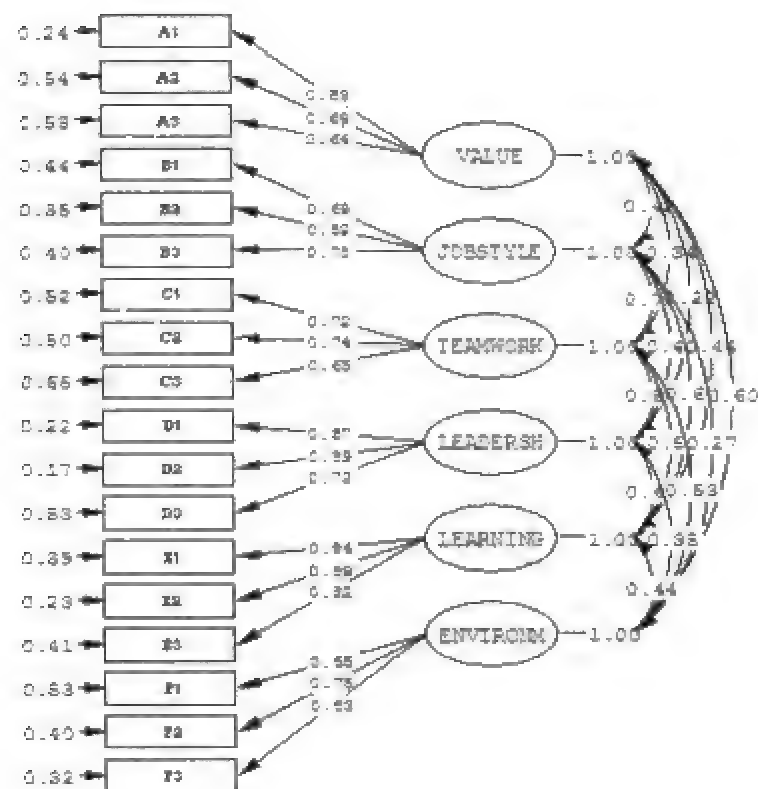
```
MO LX=IN TD=IN
```

模型四的设置, 另外增加一项 $PH=IN$ 设定, 但是要将 FIX 指令与 VALUE 指令当中有关 PH 参数的设定移除, 也就是移去表 9.1 的第 23 和第 24 行指令。MODEL 指令内容如下:

```
MO LX=IN TD=IN PH=IN
```



(a) 女性样本



(b) 男性样本

图 9.1 模型二参数估计结果

四、分析结果整理

（一）第一阶段分析结果

第一阶段系根据第五章范例语法分别对全体样本、男性样本、女性样本进行 CFA 分析。由于分析结果与第五章范例类似，细节不在此赘述，仅列出摘要表于表 9.3。

由表 9.3 数据可知，三个模型的 CFA 分析均有相同的自由度，显示模型界定与因素结构完全相等。但是，男性样本似乎有较为理想的模型拟合度，卡方值较小，且 RMSEA 指数较低。但是如果以 NNFI 与 CFI 指数来看，三者的差异不大，显示此一 CFA 模型在全体样本以及个别样本的拟合情形还算一致，因此可以进行下一个阶段的跨样本分析。

值得注意的是，由于多重样本分析的第一个模型是基准模型，两个样本间没有任何等同设定，因此表 9.3 当中的男性样本与女性样本的个别卡方值（自由度）相加，即等于第二阶段模型一的卡方值（自由度），读者可以自行验算之。

表 9.3 第一阶段单样本 CFA 分析模型拟合度评估摘要表

Model	χ^2 and WLS χ^2	df	RMSEA	NNFI	CFI	GFI	SRMR
全体样本	239.89 (P=.000) 228.06 (P=.000)	120	.054	.98	.98	.92	.052
男性样本	192.72 (P=.000) 176.20 (P=.001)	120	.052	.96	.97	.90	.064
女性样本	219.48 (P=.000) 209.15 (P=.000)	120	.073	.96	.97	.86	.066

（二）第二阶段分析结果

第二阶段的多样本模型分析，总计有四个模型进行不同程度的恒等限定，下面仅以第四个模型，也就是 Λ_x 、 Φ 、 Θ_ε 三个矩阵均被限定为样本间等同。报表内容(Ch9amg4.out)如下：

首先, 参数设定指令说明了整个模型当中, 共有 51 个被估计参数。其中被设定为相等的参数包括 Λ 矩阵、 Θ 矩阵与 Φ 矩阵 (注 1)。而为了使潜在变量能够量尺化, 各潜在变量方差设定为 1, 不予估计 (注 2)。

注 1: 两个样本的 Λ 、 Θ 、 Φ 矩阵设为相等, 因此仅有一组参数被估计。

Parameter Specifications

LAMBDA-X EQUALS LAMBDA-X IN THE FOLLOWING GROUP

PHI EQUALS PHI IN THE FOLLOWING GROUP

THETA-DELTA EQUALS THETA-DELTA IN THE FOLLOWING GROUP

GROUP: FEMALE

Parameter Specifications

LAMBDA-X

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
A1	1	0	0	0	0	0
A2	2	0	0	0	0	0
A3	3	0	0	0	0	0
B1	0	4	0	0	0	0
B2	0	5	0	0	0	0
B3	0	6	0	0	0	0
C1	0	0	7	0	0	0
C2	0	0	8	0	0	0
C3	0	0	9	0	0	0
D1	0	0	0	10	0	0
D2	0	0	0	11	0	0
D3	0	0	0	12	0	0
E1	0	0	0	0	13	0
E2	0	0	0	0	14	0
E3	0	0	0	0	15	0
F1	0	0	0	0	0	16
F2	0	0	0	0	0	17
F3	0	0	0	0	0	18

PHI

LAMBDA-X

	VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
VALUE	0					
JOBSTYLE	19	0				
TEAMWORK	20	21	0			
LEADERSH	22	23	24	0		
LEARNING	25	26	27	28	0	
ENVIRONM	29	30	31	32	33	0

注 2: 各潜在变量的方差被设定为 1, 故不予估计。

THETA-DELTA					
A1	A2	A3	B1	B2	B3
-----	-----	-----	-----	-----	-----
34	35	36	37	38	39
THETA-DELTA					
C1	C2	C3	D1	D2	D3
-----	-----	-----	-----	-----	-----
40	41	42	43	44	45
THETA-DELTA					
E1	E2	E3	F1	F2	F3
-----	-----	-----	-----	-----	-----
46	47	48	49	50	51

(三) 参数估计结果

多样本 CFA 参数估计的原理与单样本 CFA 相同，仅有的差异是多样本所处理的参数估计随着组数的增加而成倍增加。但是以模型四为例，所有的参数被设定为相等，因此，尽管样本组数为二，但是估计的参数数目与单样本 CFA 相同（均为 51 个参数）。

另外，由于多样本模型分析所着重的是当参数加以限定为等同之后，模型拟合度的变化是否有意义；因此，本范例省略参数估计结果的说明，直接就各模型的模型拟合度数据加以比较。

(四) 模型拟合度分析

模型一至模型四，加上阶段一的三个模型，总计七个模型的模型拟合度评估结果列举于表 9.4。

由表 9.4 的数据显示，所有的 CFA 模型的拟合指数都十分的接近，但是，阶段二的模型则随着恒等限制的增加，释放的自由参数越多，卡方值逐渐增加。显示恒等限制的加入对于模型估计产生负面的影响。因为恒等限制的各模型都是属于嵌套模型（nested models），因模型间的拟合度差异量可以利用卡方差异检验来考验。

模型一将两个独立 CFA 模型整合于同一个 CFA 分析当中，其性质就是在检测协方差矩阵是否等同的假设，卡方检验的结果拒绝了此一假设，因此可以继续进行更进一步的恒等性评估，检验因素效度的不稳定因素在哪里。

模型二与模型一的卡方差异量检验，增加了因素载荷的恒等限制后，模型的拟

合度降低, $\Delta\chi^2=39.28$, $\Delta df=18$ ($p<.01$), 显示两个样本在因素载荷参数上即有明显的差异, 因素载荷是不等同的。此时, 研究者应该停止进一步的恒等性检验, 就两个样本的因素载荷的数值内容进行详细的讨论。但是, 为了示范进一步的恒等性检验, 我们仍然继续讨论模型三的卡方差异量检验结果。

模型三与模型二的比较, 以及模型四与模型三的比较, 都获得一个显著的卡方差异量 ($\Delta\chi^2=38.75$, $\Delta df=18$, $p<.01$ 以及 $\Delta\chi^2=46.12$, $\Delta df=15$, $p<.01$), 显示不论是增加残差变异量的恒等假设, 或是增加因素间相关恒等假设, 都造成了显著的模型拟合度降低, 表示样本间的等同性是不存在的。

根据此一结果, 我们应退回阶段一的两个样本的单样本 CFA 检验, 探讨此一六因素的量表在男女性样本上的差异为何。尽管这个差异是不容易被清楚明确的界定, 但是在量表的学术与实用价值上, 此一程序均具有其价值。例如, 研究者可以利用这个检查程序来进行项目分析, 检讨为何题目间缺乏测量恒等性, 必要时可以将题目删除。

表 9.4 多样本 CFA 分析模型拟合度评估摘要表

Model	WLS χ^2	DF	NCP	RMSEA	NNFI	CFI	GFI
<u>阶段一</u>							
全体样本	228.06 ($p=.000$)	120	108.1	.054	.98	.98	.92
男性样本	176.20 ($p=.001$)	120	56.2	.052	.96	.97	.90
女性样本	209.15 ($p=.000$)	120	289.2	.073	.96	.97	.86
<u>阶段二</u>							
模型一	385.36 ($p=.000$)	240	145.4	.063	.96	.97	.86
模型二	424.64 ($p=.000$)	258	166.6	.065	.96	.97	.85
	$\Delta^a=39.28$	$\Delta=18$					
模型三	463.39 ($p=.000$)	276	187.4	.066	.96	.96	.84
	$\Delta=38.75$	$\Delta=18$					
模型四	509.51 ($p=.000$)	291	218.5	.070	.96	.96	.83
	$\Delta=46.12$	$\Delta=15$					

^a 差异值的计算以后一个模型减去前一个模型的数值, 例如 $39.28=424.64-385.36$; $38.75=463.39-424.64$ 。

第四节 AMOS 的多样本分析

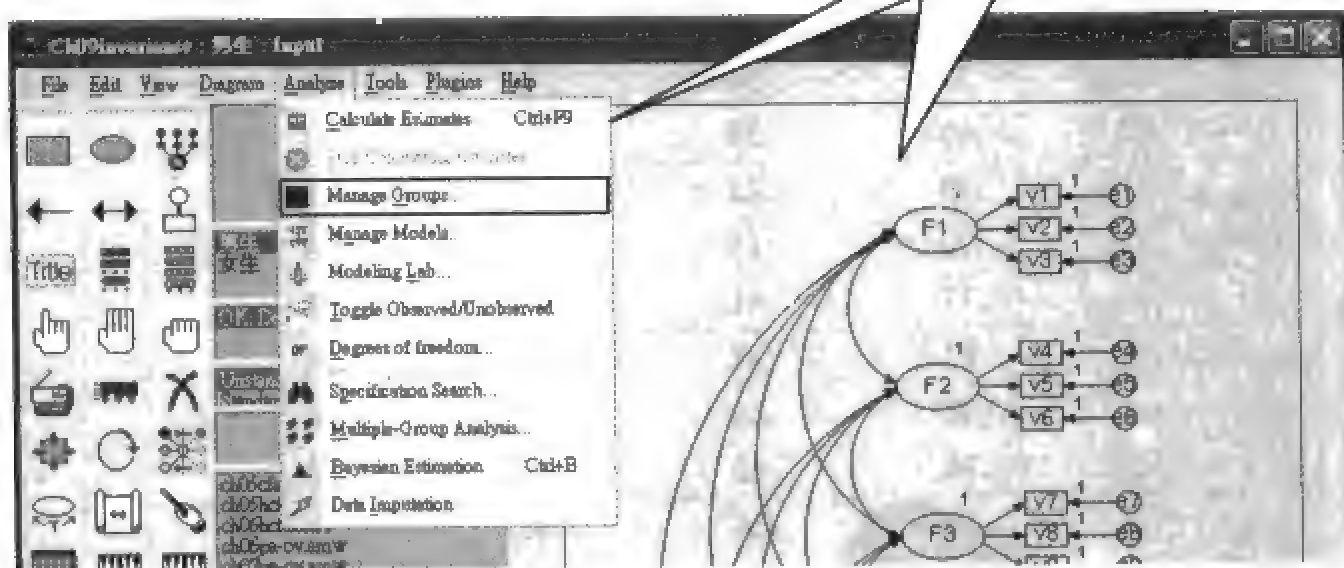
AMOS 的多样本分析操作方式非常简单, 仅需点选 AMOS 所提供的多样本分析功能, 即可利用对话框设定测量恒等性的不同模型界定。以下即针对应用多样本分析进行测量恒等性检测的特殊步骤加以说明。

一、操作步骤

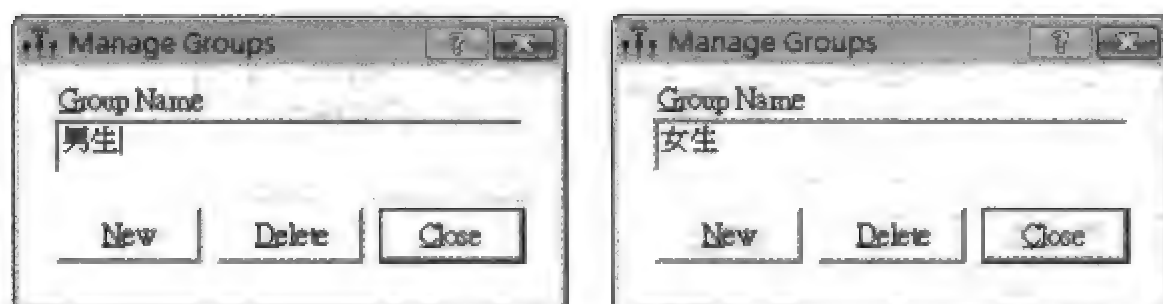
步骤一: 绘制假设模型图

步骤二: 点选 Analyze->Manage Groups

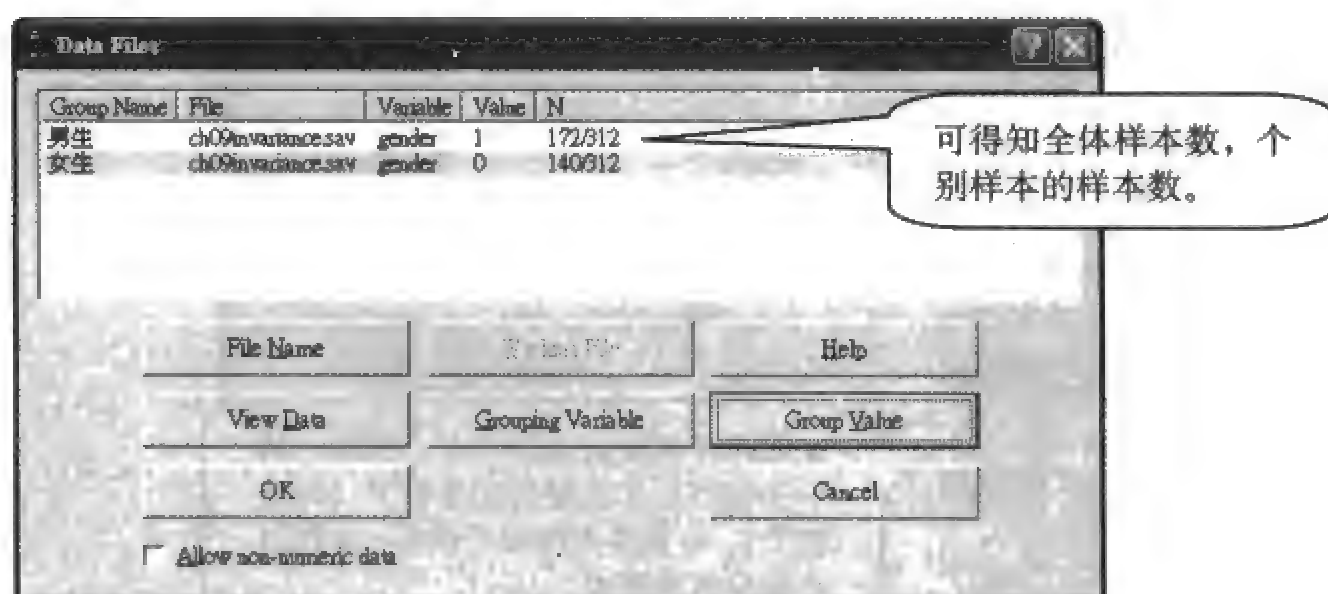
绘制假设模型图并
点选 Analyze 中的
Manage Groups



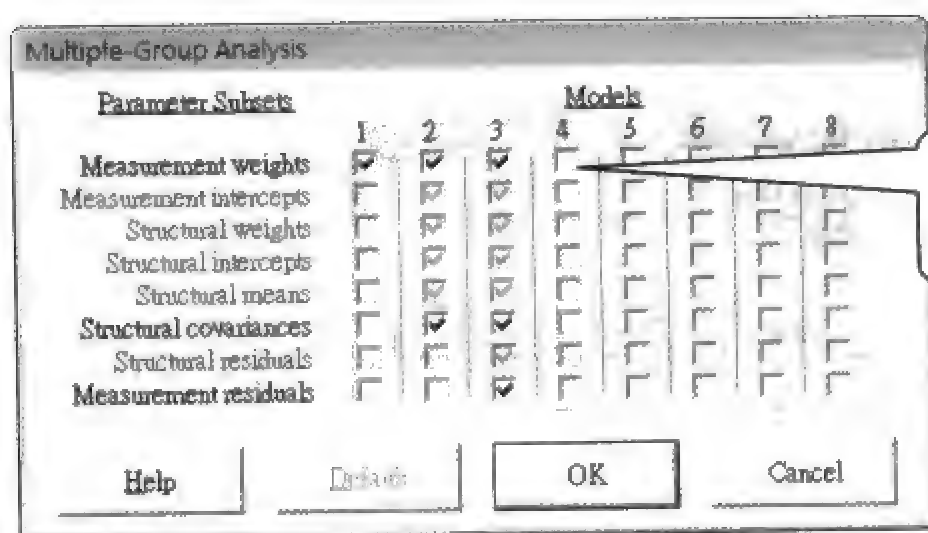
步骤三: 将 Group number1 设为 male (男生) ->点选 New。再将 Group number2 改为 Female (女生) ->点选 Close。完成分组名称设定。



步骤四：点选数据文件，开启 ch09.sav 的数据库，点选 Grouping Variable->gender，再点选 Group Value->分别选择 1 与 0（Gender 中的男生与女生分别为 1 与 0），完成后的画面如下，并可得到两者的样本数分别为 172 与 140：

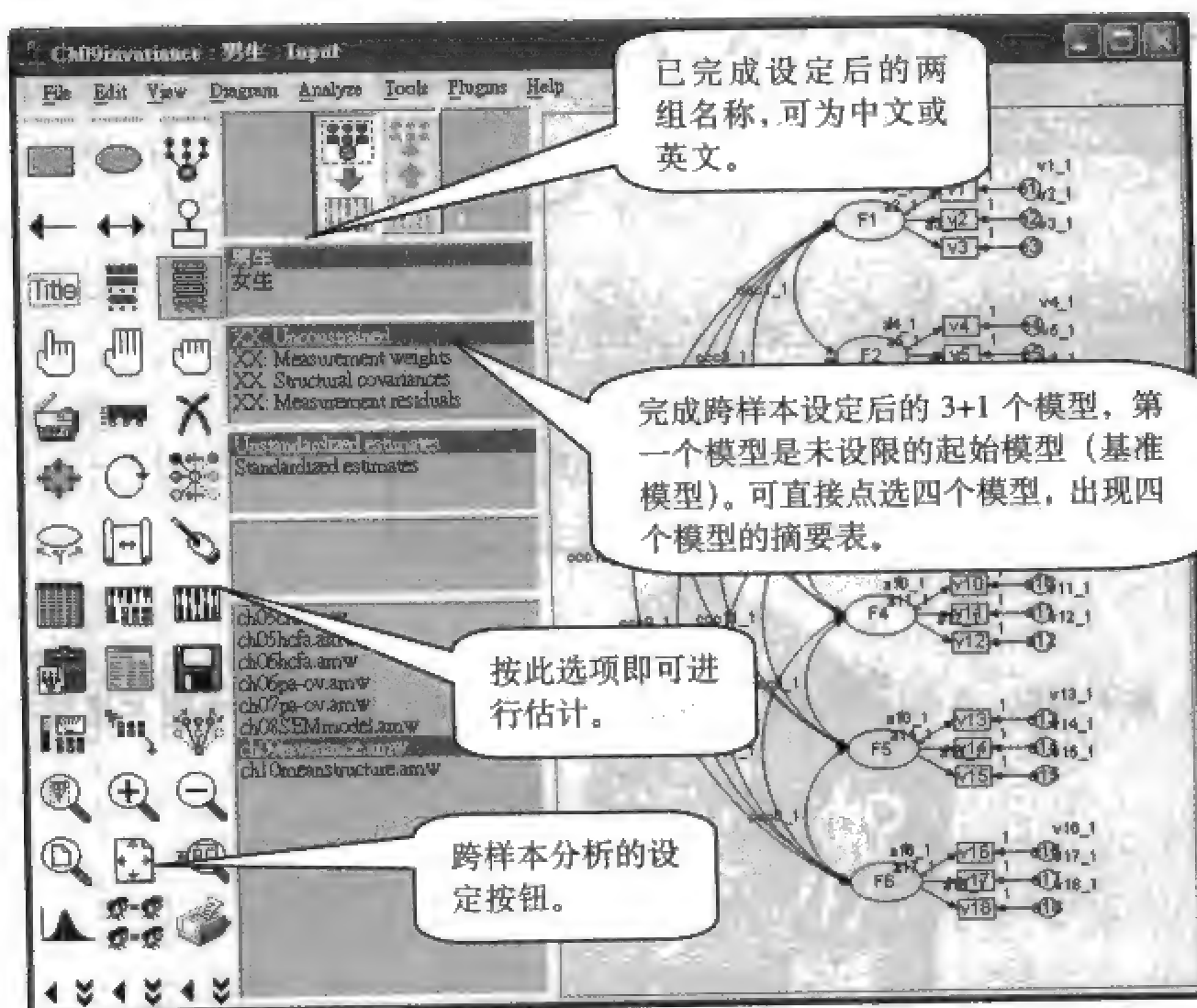


步骤五：按下 OK，点选  (多样本分析)，按确定之后即进入模型界定对话框如下：

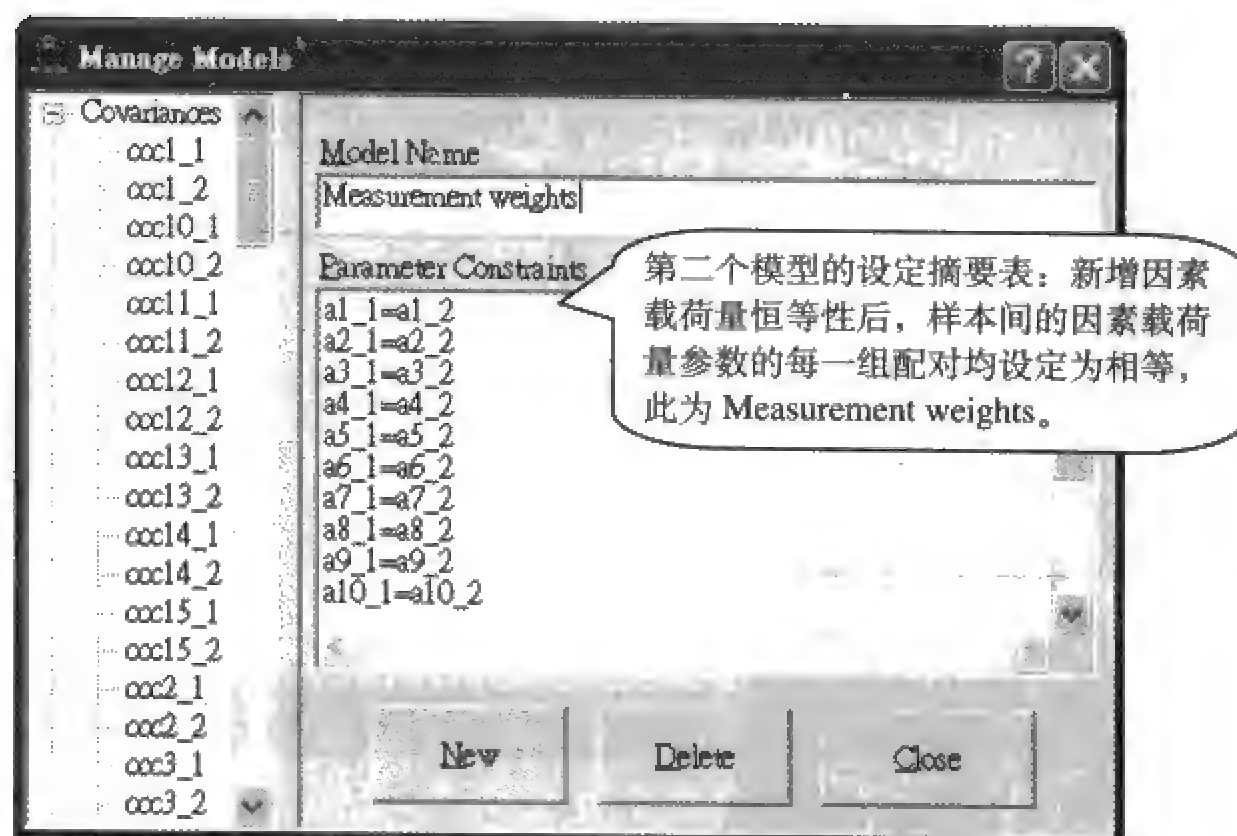
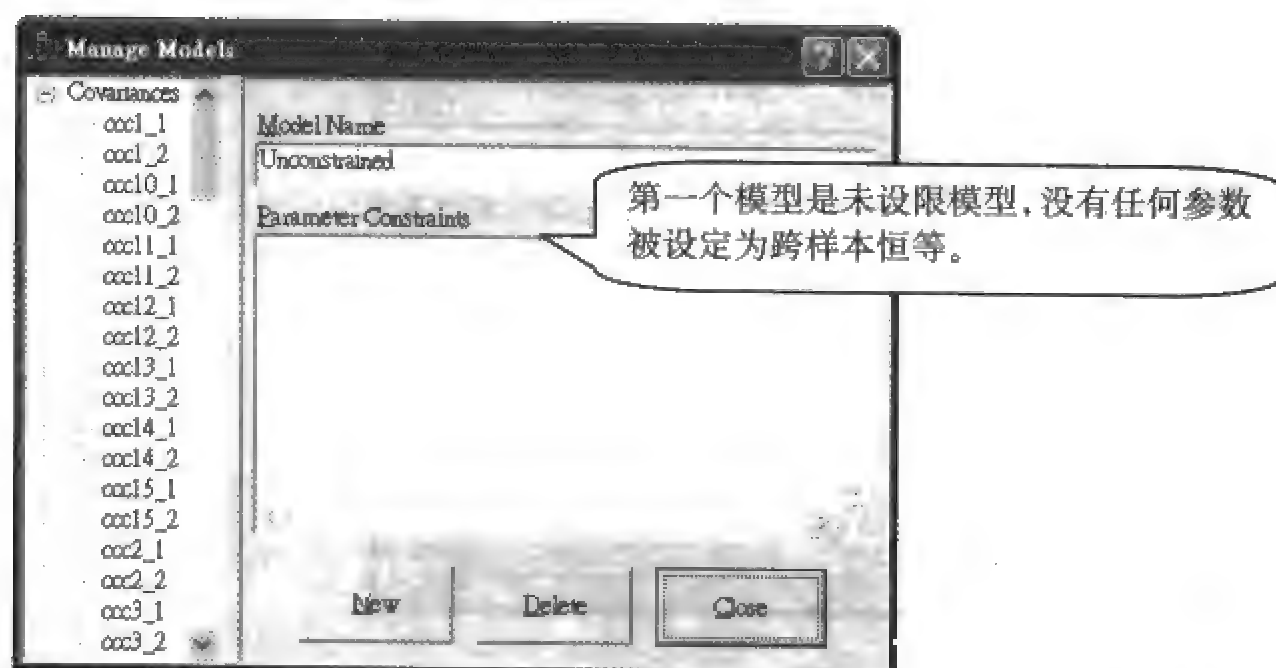


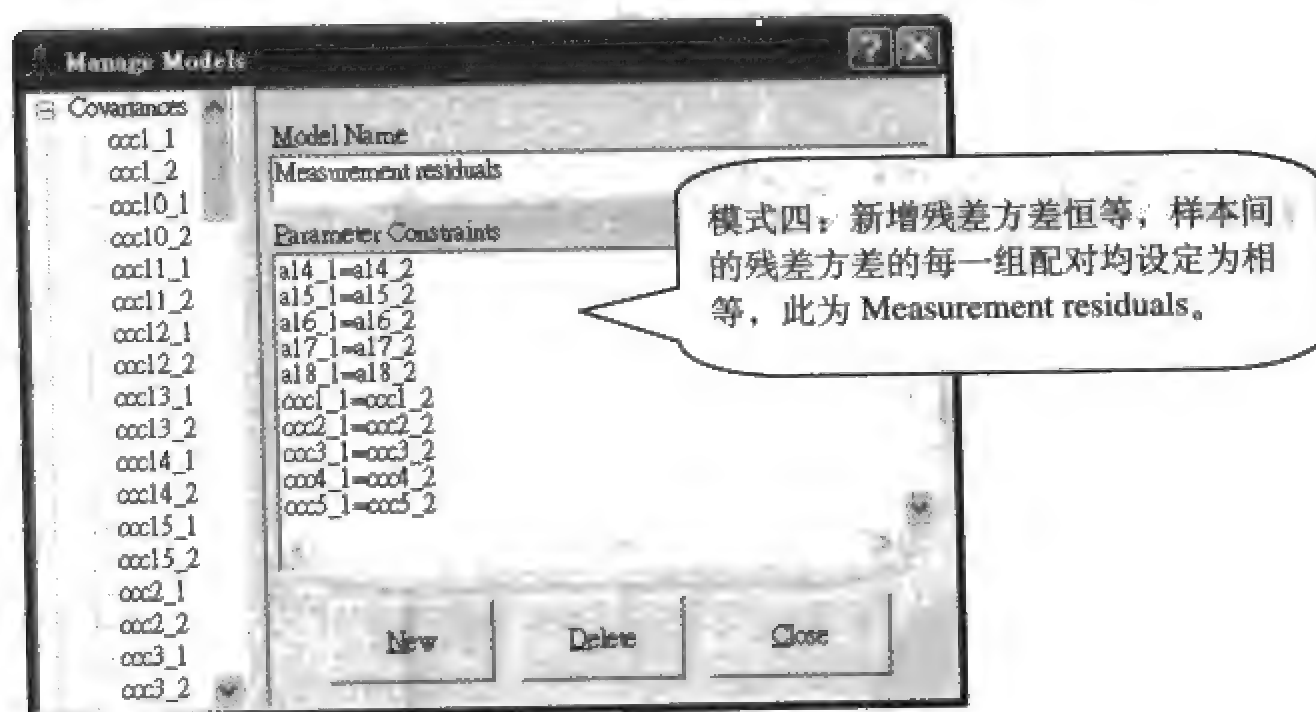
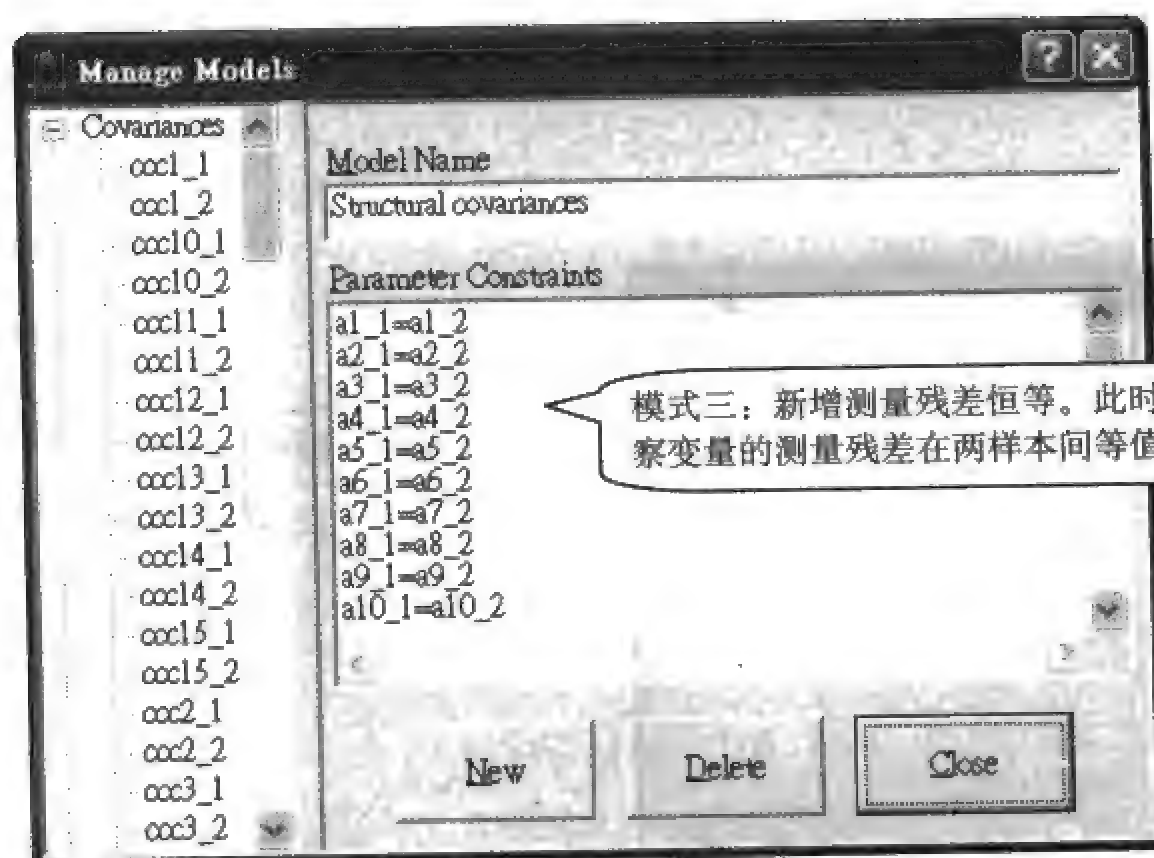
如果要增加模型，可以自行在后续模型中加以设定。

对话框中显示 AMOS 所自动默认的各种模型。打勾者表示两个样本设定等值。其中第一个模型表示因素载荷跨样本等值。第二个模型表示增加了结构协方差（亦即因素间相关）跨样本等值。第三个模型表示还增加了测量残差跨样本等值，此为完全等值模型。



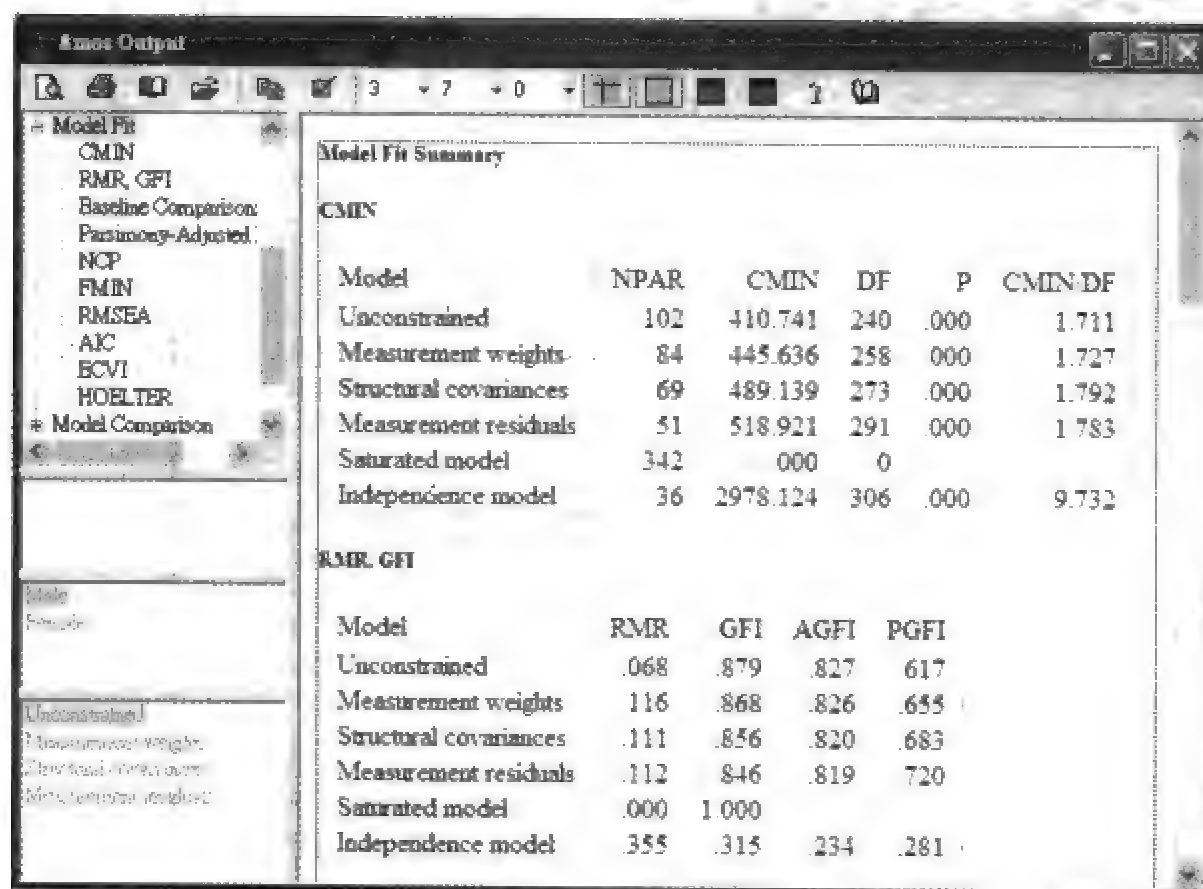
若点选 AMOS 左侧菜单中四个模型，可以得到以各模型界定摘要表（第一个模型为未设限模型）：





二、估计结果

进行估计后，AMOS 将会将四个模型的估计结果列于报表中(详细报表请参阅 ch09.AmosOutput)。以下是模型拟合指数：



The screenshot shows the AMOS Output window with the following content:

Model Fit Summary

CMIN

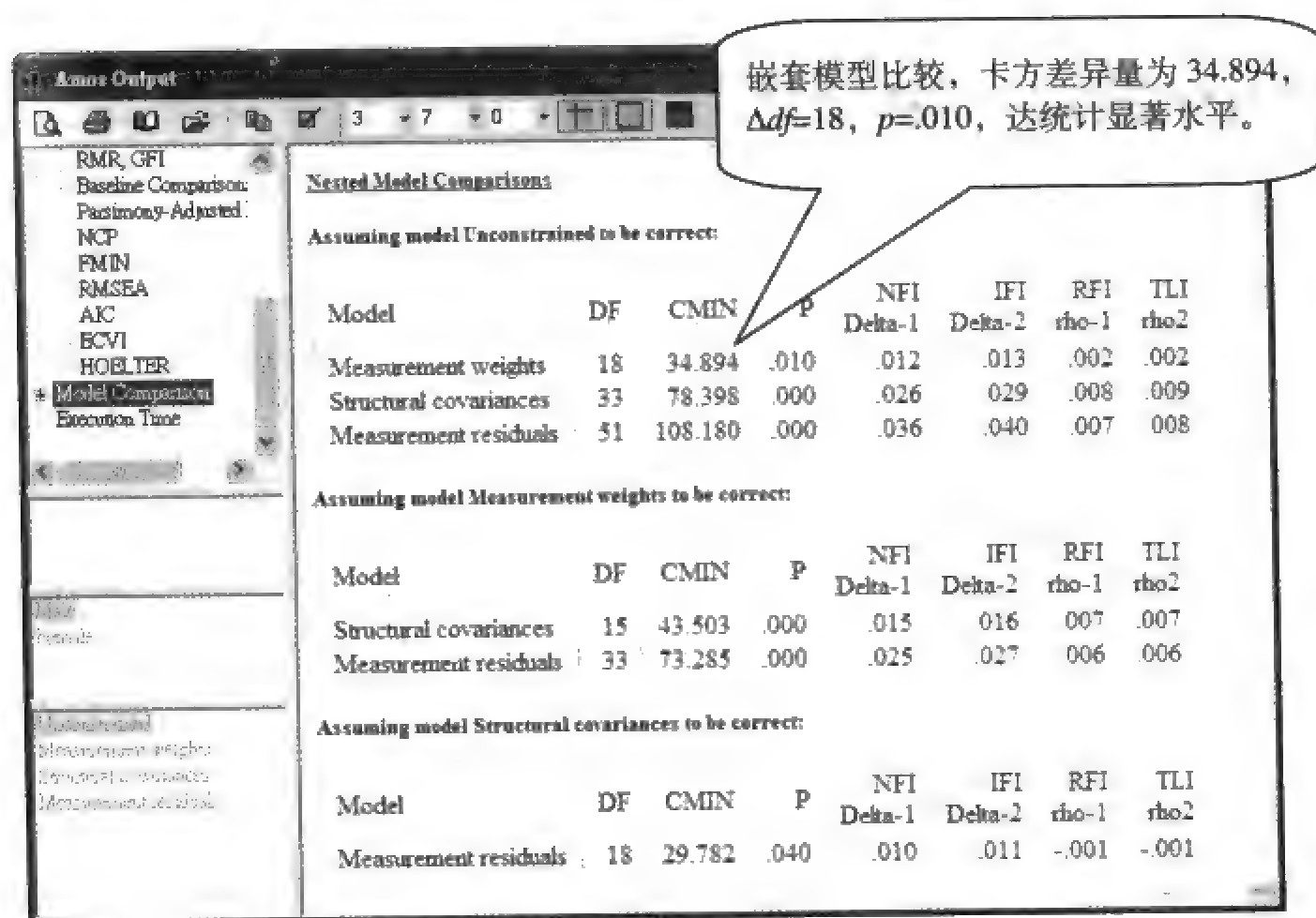
Model	NPAR	CMIN	DF	P	CMIN/DF
Unconstrained	102	410.741	240	.000	1.711
Measurement weights	84	445.636	258	.000	1.727
Structural covariances	69	489.139	273	.000	1.792
Measurement residuals	51	518.921	291	.000	1.783
Saturated model	342	.000	0		
Independence model	36	2978.124	306	.000	9.732

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Unconstrained	.068	.879	.827	.617
Measurement weights	.116	.868	.826	.635
Structural covariances	.111	.856	.820	.683
Measurement residuals	.112	.846	.819	.720
Saturated model	.000	1.000		
Independence model	.355	.315	.234	.281

由于四个模型各有不同的恒等限制，属于嵌套模型，模型间比较可利用卡方差异检验来检验。AMOS 也已经在报表中整理出这些模型以第一个未设限模型为基准所进行的嵌套比较（卡方差异检验）结果。

以因素载荷设限的模型为例，卡方差异值为 34.894 ($p=.010$) 达统计显著水平，显示因素载荷设定为恒等后，模型拟合显著变差，而 34.894 的数值是 measurement weight 模型的卡方值 (445.636) 与未设限模型卡方值 (410.741) 的差异值。



对于参数估计的结果，则可以在报表的左侧选单中，选择四种模型在两个样本上的估计值。如果设定为等值者，则在两个样本上的参数估计得到相同的数值。

以第一个设限模型（Measurement Weight 模型）为例，两个样本在协方差估计值应各有数值，不会等值：

Amos Output

3 7 0

Variable Summary
Parameter summary
Notes for Model
Estimates
Scalars
Regression Weights
Standardized Regression Weights
Covariances
Correlations
Variances
Squared Multiple Correlations
Minimization History

Male
Female

Unconstrained
Measurement weights
Structural covariances
Measurement residuals

Covariances: (Male - Measurement weights)

	Estimate	S.E.	C.R.	P	Label
F1 <--> F2	.640	.065	9.815	***	ccc1_1
F2 <--> F3	.634	.067	9.519	***	ccc2_1
F1 <--> F3	.626	.070	8.909	***	ccc3_1
F3 <--> F4	.492	.074	6.670	***	ccc4_1
F4 <--> F5	.533	.062	8.604	***	ccc5_1
F5 <--> F6	.389	.077	5.082	***	ccc6_1
F4 <--> F6	.316	.081	3.883	***	ccc7_1
F3 <--> F6	.652	.068	9.651	***	ccc8_1
F2 <--> F6	.457	.078	5.893	***	ccc9_1
F1 <--> F6	.760	.057	13.258	***	ccc10_1
F3 <--> F5	.609	.063	9.651	***	ccc11_1
F2 <--> F5	.529	.066	8.067	***	ccc12_1
F1 <--> F5	.585	.064	9.120	***	ccc13_1
F2 <--> F4	.436	.074	5.915	***	ccc14_1
F1 <--> F4	.661	.060	11.112	***	ccc15_1

协方差估计值, 标准误与显著性检验。这些估计值并未设定跨样本恒等, 因而会与女性样本不同值。

选择第一个设限模型, 第一个样本 (Male) 的协方差估计值。

Amos Output

3 7 0

Variable Summary
Parameter summary
Notes for Model
Estimates
Scalars
Regression Weights
Standardized Regression Weights
Covariances
Correlations
Variances
Squared Multiple Correlations
Minimization History

Male
Female

Unconstrained
Measurement weights
Structural covariances
Measurement residuals

Covariances: (Female - Measurement weights)

	Estimate	S.E.	C.R.	P	Label
F1 <--> F2	.420	.088	4.803	***	ccc1_2
F2 <--> F3	.781	.062	12.625	***	ccc2_2
F1 <--> F3	.339	.098	3.456	***	ccc3_2
F3 <--> F4	.572	.076	7.534	***	ccc4_2
F4 <--> F5	.610	.062	9.878	***	ccc5_2
F5 <--> F6	.446	.083	5.370	***	ccc6_2
F4 <--> F6	.360	.087	4.142	***	ccc7_2
F3 <--> F6	.536	.087	6.183	***	ccc8_2
F2 <--> F6	.273	.095	2.858	.004	ccc9_2
F1 <--> F6	.593	.075	7.857	***	ccc10_2
F3 <--> F5	.582	.077	7.602	***	ccc11_2
F2 <--> F5	.617	.067	9.271	***	ccc12_2
F1 <--> F5	.428	.083	5.134	***	ccc13_2
F2 <--> F4	.6394	.074	8.6394	***	ccc14_2
F1 <--> F4	.2340	.019	12.340	***	ccc15_2

第一个设限模型的第二个样本 (Female) 的协方差估计值, 与 Male 不同。

第五节 Mplus 的多样本分析

以 Mplus 执行测量恒等性分析的语法 (以第四个模型, 也就是 Λ 、 Φ 、 Θ 三个矩阵均被限定为样本间等同为例)。

一、Mplus 语法

TITLE: Ch9a Multi-group CFA Model 4

DATA:

```
NGROUPS=2;
FILE (male) IS Ch9am.dat;
FILE (female) IS Ch9af.dat;
```

多样本分析的设定格式

NGGROUP=2 表示有两个样本,
数据各自存放于两个档案中。

VARIABLE: NAMES ARE a1-a3 b1-b3 c1-c3 d1-d3 e1-e3 f1-f3;

MODEL:

```
VALUE by a1*a2-a3;
JOBSTYLE by b1*b2-b3;
TEAMWORK by c1*c2-c3;
LEADERSH by d1*d2-d3;
LEARNING by e1*e2-e3;
ENVIRONM by f1*f2-f3;
```

设定共变为跨样本恒等

如果这些语法移除, 则为模型三。

```
VALUE with JOBSTYLE (19);
VALUE with TEAMWORK (20);
VALUE with LEADERSH (21);
VALUE with LEARNING (22);
VALUE with ENVIRONM (23);
JOBSTYLE with TEAMWORK (24);
JOBSTYLE with LEADERSH (25);
JOBSTYLE with LEARNING (26);
JOBSTYLE with ENVIRONM (27);
TEAMWORK with LEADERSH (28);
TEAMWORK with LEARNING (29);
TEAMWORK with ENVIRONM (30);
LEADERSH with LEARNING (31);
LEADERSH with ENVIRONM (32);
LEARNING with ENVIRONM (33);
```

设定观察变量残差具跨样本恒等, 模型
一无此项设定。

观察变量截距在两样本间
自由估计。

```
a1(1); a2(2); a3(3); b1(4); b2(5); b3(6); c1(7); c2(8); c3(9); d1(10);
d2(11); d3(12); e1(13); e2(14); e3(15); f1(16); f2(17); f3(18);
```

```
VALUE@1 JOBSTYLE@1 TEAMWORK@1 LEADERSH@1 LEARNING@1 ENVIRONM@1;
[a1 a2 a3 b1 b2 b3 c1 c2 c3 d1 d2 d3 e1 e2 e3 f1 f2 f3];
[VALUE@0 JOBSTYLE@0 TEAMWORK@0 LEADERSH@0 LEARNING@0 ENVIRONM@0];
```

MODEL male:

```
[a1 a2 a3 b1 b2 b3 c1 c2 c3 d1 d2 d3 e1 e2 e3 f1 f2 f3];
```

MODEL female:

```
[a1 a2 a3 b1 b2 b3 c1 c2 c3 d1 d2 d3 e1 e2 e3 f1 f2 f3];
```

二、Mplus 报表

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value	522.269
Degrees of Freedom	291
P-Value	0.0000

两个样本各自估计的模适配指针卡方值，合计即为总和的 522.269。

Chi-Square Contributions From Each Group

MALE	251.685
FEMALE	270.584

Chi-Square Test of Model Fit for the Baseline Model

Value	2997.338
Degrees of Freedom	306
P-Value	0.0000

CFI/TLI

CFI	0.914
TLI	0.910

Loglikelihood

H0 Value	-6708.937
H1 Value	-6447.803

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.071
90 Percent C.I.	0.061 0.081

SRMR (Standardized Root Mean Square Residual)

Value	0.127
-------	-------

MODEL RESULTS

以下为男性样本估计数

Group	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
MALE				
VALUE BY				
A1	0.795	0.052	15.329	0.000
A2	0.691	0.056	12.311	0.000
A3	0.603	0.054	11.197	0.000
JOBSTYLE BY				
B1	0.784	0.062	12.632	0.000
B2	0.941	0.058	16.266	0.000
B3	0.833	0.055	15.063	0.000

TEAMWORK BY				
C1	0.690	0.054	12.884	0.000
C2	0.749	0.056	13.259	0.000
..... (略)				
VALUE WITH				
JOBSTYLE	0.520	0.057	9.074	0.000
TEAMWORK	0.475	0.064	7.377	0.000
LEADERSH	0.419	0.059	7.045	0.000
LEARNING	0.507	0.056	9.038	0.000
ENVIRONM	0.677	0.048	14.070	0.000
..... (略)				
Intercepts				
A1	4.552	0.074	61.740	0.000
A2	4.442	0.077	57.626	0.000
..... (略)				
Residual Variances				
A1	0.303	0.049	6.215	0.000
A2	0.544	0.056	9.732	0.000
A3	0.558	0.053	10.615	0.000
..... (略)				
Group FEMALE				
VALUE BY				
A1	0.795	0.052	15.329	0.000
A2	0.691	0.056	12.311	0.000
A3	0.603	0.054	11.197	0.000
..... (略)				
VALUE WITH				
JOBSTYLE	0.520	0.057	9.074	0.000
TEAMWORK	0.475	0.064	7.377	0.000
LEADERSH	0.419	0.059	7.045	0.000
LEARNING	0.507	0.056	9.038	0.000
ENVIRONM	0.677	0.048	14.070	0.000
..... (略)				
Intercepts				
A1	4.236	0.082	51.828	0.000
A2	4.136	0.085	48.407	0.000
..... (略)				
Residual Variances				
A1	0.303	0.049	6.215	0.000
A2	0.544	0.056	9.732	0.000
A3	0.558	0.053	10.615	0.000
..... (略)				

以下为女性样本估计数
因素载荷量、残差方差以及潜在变量协方差
均与男性样本相同。

第六节 多样本分析：复核效化检验

SEM 的复核效化检验可区分为三种方式，分别为宽松复核 (loose replication strategy)、温和复核 (moderate replication strategy)、严格复核 (tight replication strategy) (MacCallum, Roznowski, Mar, & Reith, 1994)。宽松取向的做法仅是将效度样本上的模式界定，套用测定样本的模式界定，但效度样本的参数仍进行自由估计。此一做法可以说是模型结构的复核，并没有做到数量的复核。相对之下，严格复核取向是将效度样本上的模型界定，不仅套用测定样本的模型界定，而且参数套以测定样本的估计数值，使两个样本的假设模型具有完全相等的条件，唯一的不同是样本的差异。在严格策略下，效度样本的模型估计是在一个非常严格的条件下进行的，因为每一个参数的估计数并不是从自己样本的观测数据所推估而得。如果效度样本在如此严格的条件下仍然可以获得良好的拟合度，表示测量的稳定度佳。如果效度样本并不是全部参数都被设定为与测定样本相等，某一些参数可以被自由估计，称为温和复核策略。MacCallum 等人 (1994) 称此种策略所得到的证据为部分复核效化 (partial cross-validation)。

复核效化可以视为是 SEM 模型的有效性分析。当我们运用 SEM 分析检验某一个假设模型得到一个理想的终解之后，可以利用其他样本进行跨样本的检测，证明模式的跨样本有效性。

在做法上，复核效化与测量恒等性评估十分类似，主要的差异是两者在利用多样本 SEM 分析时，所评估的模型对象有所不同。在测量恒等性检验中，各参数在不同样本之间具有不同的恒等性设限。相对之下，复核效化利用测定样本的参数估计数，作为效度样本的参数数值，然后仅评估整个模型在效度样本上的拟合度。

以下，我们即以先前的范例，应用复核效化的分析检验测量模型在跨样本间的稳定性。数据仍是来自于《组织创新气氛量表》十八题短题本在 255 位员工中实施的测量结果。由于这些员工分属两个分公司，因此我们可以将其中一个分公司的员工样本视为测定样本，再以另一个分公司的员工样本作为效度样本，进行复核效化检验。而《组织创新气氛量表》的测量模型则可以直接套用范例一的模型界定进行跨样本的检验。

一、复核效化的模型界定

本范例所输入的数据为协方差矩阵，测定样本的协方差矩阵数据存放在 Ch9b1.cov 档案，效度样本的协方差矩阵数据存放 Ch9b2.cov 中。

表 9.6 列出了用于宽松复核分析的 LISREL 与 SIMPLIS 语法。语法由两个独立部分所组成，第 1 至第 11 行用以设定测定样本的模型，第 12 至第 19 行用以设定效度样本的模型。两个部分除了样本数与数据文件的文件名不同以及效度样本的 MO 指令中具有恒等性设定之外，其余指令一样，显示两个样本的因素结构是相同的设定。

在 SIMPLIS 语法部分，三种不同复核效化策略的差异反映在第 15 行效度样本的设定条件上。当效度样本的模型界定当中没有特别列出时，表示参数套用测定样本估计值。

(一) 宽松复核策略

若要进行宽松复核策略分析，效度样本仅需维持与测定模型相同的因素结构设定，参数无需进行任何等同设定，因此 LISREL 语法在第 16 行的 MO 指令中，无需增加任何的恒等设定。

但在 SIMPLIS 语法，由于要让两个样本各自估计因素载荷、残差变异，因此在语法中需列出第 21 到第 26 行（因素载荷）、第 27 行（测量残差）与第 28 行以后的潜在变量协方差（共有 15 行）。如此才是两个样本各自估计参数且仅在因素结构上保持一致。

表 9.6 复核效化 LISREL 语法

LISREL 语法(Ch9b1.ls8)	
1	GROUP: Calibration sample
2	DATA NG=2 NI=18 NO=144
3	KM SY FI=Ch9b1.cov
4	LA; A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3
5	MODEL NX=18 NK=6 PH=SY,FR TD=DI,FR LX=FU,FI
6	LK; VALUE JOBSTYLE TEAMWORK LEADERSHIP LEARNING ENVIRONMENT
7	FREE LX(1,1) LX(2,1) LX(3,1) LX(4,2) LX(5,2) LX(6,2) LX(7,3) LX(8,3) LX(9,3)

```

8  FREE  LX(10,4) LX(11,4) LX(12,4) LX(13,5) LX(14,5) LX(15,5)
    LX(16,6)LX(17,6) LX(18,6)
9  VALUE 1 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
10 PD
11 OUTPUT SS SC
12 GROUP: Validation sample
13 DATA NO=111
14 KM SY FI=Ch9b2.cov
15 LA; A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3
16 MODEL LX=IN TD=IN PH=IN
17 LK; VALUE JOBSTYLE TEAMWORK LEADERSHIP LEARNING ENVIRONMENT
18 FIX PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
19 VALUE 1 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
20 PD
21 OUTPUT SS SC

```

SIMPLIS 语法(Ch9b1.spl)

```

1  GROUP: calibration sample
2  Observed variables: A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3
3  correlation matrix from file Ch9b1.cov
4  Sample Size = 144
5  Latent Variables  VALUE JOBSTYLE TEAMWORK LEADERSH LEARNING ENVIRONM
6  Relationships
7  A1-A3= VALUE
8  B1-B3 = JOBSTYLE
9  C1-C3 = TEAMWORK
10 D1-D3 = LEADERSH
11 E1-E3 = LEARNING
12 F1 -F3= ENVIRONM
13 Set the Variance of VALUE-ENVIRONM to 1.00
14 Path Diagram
15 GROUP: Validation sample
16 Observed variables: A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3
17 correlation matrix from file Ch9b2.cov
18 Sample Size = 111
19 Latent Variables  VALUE JOBSTYLE TEAMWORK LEADERSH LEARNING ENVIRONM
20 Relationships
21 A1-A3= VALUE
22 B1-B3 = JOBSTYLE
23 C1-C3 = TEAMWORK
24 D1-D3 = LEADERSH
25 E1-E3 = LEARNING

```

```

26  F1 -F3= ENVIRONM
27  Set the Error Variance of A1-F3 Free
28  Set the Covariances of JOBSTYLE and VALUE Free
29  Set the Covariances of TEAMWORK and VALUE Free
30  Set the Covariances of TEAMWORK and JOBSTYLE Free
    .....(略)
31  End of Problem

```

(二) 温和复核策略

温和复核策略的分析,是指效度样本的部分参数必须套用测定样本的参数数据。也就是说效度模型中的部分参数必须设定为样本间等同,一般惯用的等同参数是因素载荷。因此在 LISREL 的语法上,必须仿照测量恒等性模型二,在效度样本的 MO 指令中加入 LX=IN 的设定, (Λ_x 矩阵相等) 的指令,然后移除 FR 与 FI 指令中对于 LX 参数的设定即可。

在 SIMPLIS 语法,两个样本的因素载荷设为等值,仅需将第 21 到第 26 行的效度样本的因素载荷移除即可,表示效度样本的模型中的因素载荷将套用测定样本的数值。

(三) 严格复核策略

严格复核策略是指两个样本之间具有完全相等的模型界定,同时参数的数值也完全相等,因此必须在 LISREL 语法中将所有参数设定为等同。此时效度样本的 MO 指令中加入 LX=IN TD=IN PH=IN (第 16 行) 设定,而将测定样本的 FIX 指令与 VALUE 指令当中有关 PH 参数的设定移除 (语法文件为 Ch9bcv_t.ls8)。

SIMPLIS 语法则更简单,除了移除第 21 到第 26 行的因素载荷,还需移除第 27 行以后的方差与协方差设定,使效度样本完全套用测定样本的估计数。

二、单样本 CFA 分析

首先,为了使读者明了两个样本的特性,我们先以范例一所示范的 CFA 估计程序分别针对两个样本进行单样本的 CFA 检验,所得到的模型拟合度数据如表 9.7。

表 9.7 单样本 CFA 分析模型拟合度评估摘要表

χ^2		RMSEA	NNFI	CFI	ECVI (90%置信区间)
MFF ^a χ^2	WLS ^b χ^2				
测定样本					
194.49	178.02	.058	.96	.97	1.96
(df=120, <i>p</i> =.000)					(1.74-2.24)
效度样本					
240.03	222.05	.088	.94	.95	2.95
(df=120, <i>p</i> =.000)					(2.60-3.36)

^a MFF 值为 Minimum Fit Function Chi-Square 的 χ^2 值。
^b WLS 值为 Normal Theory Weighted Least Squares Chi-Square 的 χ^2 值。

由表 9.7 数据可知，两个模型的 CFA 分析均有相同的自由度，显示模型界定与因素结构完全相等。测定样本的拟合度显示模型的拟合度尚佳，WLS $\chi^2_{(120)}=178.02$ ($p=.000$)， $RMSEA=.058$ ， $NNFI=.96$ ， $CFI=.97$ ，效度样本的拟合度则显得较为不理想，WLS $\chi^2_{(120)}=222.05$ ($p=.000$)， $RMSEA=.088$ ， $NNFI=.94$ ， $CFI=.95$ 。样本间的复核效化检验需进一步利用跨样本分析进行。

三、复核效化检验结果

复核效化的检验主要在检查卡方贡献统计量 (contribution to Chi-square) 与卡方贡献百分比 (percentage contribution to Chi-square)。这两项数据代表了效度样本的加入后，对整体模型拟合所造成的影响。

为了使读者明了复核效化数据的意义，以下所进行的复核效化检验，除了以原先设定的测定样本与效度样本进行多样本分析，并将两个样本的角色倒置，亦即以效度样本为测定样本，以测定样本为效度样本，得到反向的复核效化评估数据以比较两者分析结果的差异。但是在实际研究上，使用者并不需要进行倒置的复核效化检验。

(一) 宽松复核策略

跨样本 CFA 模型估计结果显示, 整体模型的拟合度尚称理想, 整体模型的自由度为 240, 卡方值为 400.06 ($p = .00$), 以宽松复核策略所针对效度样本加入后的卡方贡献统计量为 240.03, 占了整体模型卡方值的 55.24% (卡方贡献百分比)。也就是说, 效度样本的 CFA 模型估计所产生的卡方值占 55.24%, 测定样本的 CFA 模型估计所产生的卡方值则为 194.49, 占 44.76%, 两者相差 11.52%。报表数据 (Ch09bcv_1.out) 如下:

```

Degrees of Freedom = 240
Minimum Fit Function Chi-Square = 434.52 (P = 0.00)
Normal Theory Weighted Least Squares Chi-Square = 400.06 (P = 0.00)
Estimated Non-centrality Parameter (NCP) = 160.06
90 Percent Confidence Interval for NCP = (108.92 ; 219.10)

Expected Cross-Validation Index (ECVI) = 2.39
90 Percent Confidence Interval for ECVI = (2.19 ; 2.62)
ECVI for Saturated Model = 1.35
ECVI for Independence Model = 21.20

Contribution to Chi-Square = 240.03
Percentage Contribution to Chi-Square = 55.24

```

若将两个样本的角色互换进行倒置宽松复核效化检验, 以原本的测定样本作为效度样本, 加入后的卡方贡献统计量为 194.49, 占了整体模型卡方值的 44.76%, 报表数据 (Ch09bcv_lr.out) 如下所示:

```

Degrees of Freedom = 240
Minimum Fit Function Chi-Square = 434.52 (P = 0.00)
Normal Theory Weighted Least Squares Chi-Square = 400.06 (P = 0.00)

Expected Cross-Validation Index (ECVI) = 2.39
90 Percent Confidence Interval for ECVI = (2.19 ; 2.62)
ECVI for Saturated Model = 1.35
ECVI for Independence Model = 21.20
Contribution to Chi-Square = 194.49
Percentage Contribution to Chi-Square = 44.76

```

由这两个程序来看, 效度样本加入后的卡方贡献统计量分别为 240.03 与 194.49,

这两个数值也就是两个样本分别进行单一样本 CFA 分析的拟合函数卡方值 (Minimum Fit Function Chi-Square) (见表 9.7)。其总和为 434.52, 也恰好为多样本分析中的拟合函数的卡方值。百分比加总为 100%, 显示两种程序只是样本角色的对调, 数学估计程序则完全相同。

(二) 温和复核策略

温和复核策略是以因素载荷设定为等同, 使两个样本具有部分等同的特性。跨样本 CFA 模型估计结果显示, 整体模型自由度为 258, 卡方值为 443.31 ($p = .00$), 效度样本加入后的卡方贡献统计量为 215.83, 占了整体模型卡方值的 45.16% (卡方贡献百分比) (Ch09bcv_m.out)。倒置策略则显示测定样本作为效度样本的卡方贡献值为 262.05, 占 54.84%。报表数据可自 Ch09bcv_m.out 与 Ch09bcv_mr.out 档案中获得。

(三) 严格复核策略

当所有参数均设定为等同的跨样本 CFA 模型估计结果显示, 整体模型自由度为 291, 卡方值为 555.02 ($p = .00$), 效度样本加入后的卡方贡献统计量为 252.91, 占了整体模型卡方值的 45.57% (卡方贡献百分比) (Ch09bcv_t.out)。相对之下, 测定样本的 CFA 模型估计所产生的卡方值为 54.43%。原始程序与倒置程序的报表数据存放于 Ch09bcv_t.out 与 Ch09bcv_tr.out。

值得注意的是, 在温和与严格复核策略中, 效度样本增加后对于整体拟合的增加量 MFF 卡方值, 已经不是个别样本进行 CFA 分析的 MFF 卡方值, 数值会较个别样本 CFA 分析的 MFF 为大。例如表 9.8 中温和策略的效度样本 MFF 卡方值为 262.05, 较宽松策略的 240.03 为大; 而严格策略的 MFF 卡方值 (302.11) 又较温和策略的 262.05 为大。显示恒等设定改变了效度样本模型拟合度。

(四) 复核效化的证据

从上述三种策略所得到的数据, 可以知道测定样本的 CFA 模型套用到效度样本后卡方值增加, 显示测量模型套用到效度模型后, 模型拟合度似乎降低了。以严格

策略为例，效度样本的所有参数均套用了测定样本导出的参数数值，此时效度样本的 MFF 卡方值为 302.11，较个别样本 CFA 的 MFF 卡方值（240.03）增加了 62.08。但是，由于严格策略的效度样本参数估计数，较单一样本 CFA 模型少了 51 个参数，因此，若将此卡方增量 $\Delta\chi^2$ 进行卡方差异检验，以 51 为自由度，显著水平为.05 时，临界值为 67.50。由于卡方差异检验 62.08 小于 67.5 的临界值，未达.05 显著水平，表示卡方增量是一种随机变化量。结果显示，从测定样本套用到效度样本的复核效化（稳定度）是被支持的。

表 9.8 不同复核效化评估策略的拟合度评估摘要表

Overall Model fit			Validity sample		
MFF $\chi^2(df)$	WLS χ^2	ECVI		MFF χ^2	%
宽松策略					
434.52(240)	400.06	2.39	原始	240.03	55.24
			倒置	194.49	44.76
温和策略					
477.88(258)	443.31	2.42	原始	262.05	54.84
			倒置	215.83	45.16
$\Delta\chi^2=77.82^a, \Delta df=18, p<.01$			$\Delta\chi^2=22.02^b, \Delta df=18, p>.05$		
			$\Delta\chi^2=21.34^b, \Delta df=18, p>.05$		
严格策略					
555.02(291)	531.45	2.50	原始	302.11	54.43
			倒置	252.91	45.57
$\Delta\chi^2=77.14^a, \Delta df=33, p<.01$			$\Delta\chi^2=62.08^b, \Delta df=51, p>.05$		
			$\Delta\chi^2=58.42^b, \Delta df=51, p>.05$		

^a 差异值的计算是以后一个复核策略减去前一个策略的 overall WLS χ^2 值。

^b 差异值的计算是以后一个复核策略减去宽松策略的效度样本的 MFF χ^2 值。原始程序为 240.03，倒置程序为 194.49。

再以温和策略为例，效度样本的因素载荷套用了测定样本的载荷后，效度样本的卡方贡献值为 262.05，较效度样本的单样本 CFA 卡方值（240.03）增加了 22.02

($\Delta\chi^2$)。此卡方增量的卡方差异检验临界值为 28.87, 亦未达 .05 显著水平, 表示从测定样本套用到效度样本的部分复核效化亦被支持。如果以倒置程序进行复核效化的检验, 效度样本的卡方增量仍未到达显著水平 (参考表 9.8), 因此, 整体来说, 本范例的复核效化检验结果支持了测量的复核效化的存在。

四、不同复核效化检验策略的比较

由上述三种复核效化估计程序可以看出, 当复核效化的限制条件越趋于严格时, 效度模型的拟合度越来越不理想。由于这三种程序的模式界定具有嵌套关系: 严格模型的设定较温和模型增多了两项恒等性设定, 而温和模型与宽松模型增加了因素载荷的恒等设定, 因此, 三种策略的估计结果可以利用卡方差异检验来比较。

由表 9.8 的数据可知, 温和与严格策略模型拟合度的卡方增量均达显著水平, 表示越趋严格的复核效化检验, 整体模型的拟合度越差, 测量稳定性越低。也就是说, 利用越严格的复核检验程序, 对于复核效化的检验越趋于严格。

从 ECVI 指数来看, 由于越小的 ECVI 值表示模型的复核效化越高, 因此表 9.8 的数据显示, 随着严格程度数值逐渐增加, 三种策略的 ECVI 指数逐渐增加, 表示复核效化越来越不理想。但是三个 ECVI 值都落在彼此的 95% 置信区间内, 表示 ECVI 值的变化在随机变化的范围内, 三种复核效化估计程序的 ECVI 值不能说有显著的不同。

从这些数据我们可以得知, 在进行复核效化检验时, 使用越严格的程序对于稳定性的检验, 条件越是严苛。但是, 也就是在越严苛的条件下, 如果效度样本的卡方增量未达显著, 反映了复核效化的证据越强。因此, 建议使用者采用最严格的程序检验复核效化, 但是如果效度样本的卡方增量达到显著, 可以改采温和的策略, 可能会得到效度样本卡方增量不显著的结果。此时证明了部分复核效化的存在, 但是相对于严格策略来说, 以温和策略得到的复核效化证据是相对薄弱的。

另外, 值得一提的是, 以宽松策略进行复核效化检验, 其原理及分析结果皆与测量恒等性检验当中的模型一 (基准模型) 检验完全相同。在基准模型中, 跨样本之间没有任何恒等假设, 仅因素结构假设相等, 也就是两个独立无关联但结构相同的 CFA

模型的组合,其卡方值为两个个别样本以同一个因素结构进行估计的总和。此时是在检验两个样本的因素结构是否等同,对于测量的复核效化的证据最为微弱。换句话说,宽松策略的复核效化检验,仅可作为复核效化的基准模型,若要符合复核效化的定义,也就是提出测量稳定性的检验证据,应采用温和策略,甚至是严格策略,以反映出测定样本作为基准模型下,再套用到效度样本的稳定程度。

整体来说,根据本范例的数据,测量模型在测定样本与效度样本间,复核效化检验的情况是被支持的,显示同一个测量模型在两个样本具有相当的稳定性,复核效化理想。而从个别样本的 CFA 分析可以看出,效度样本的拟合度较测定样本不理想,在温和与严格策略数据中也可以看出效度样本的卡方贡献量稳定的高于 50%,显示理论模型与观察数据的差距在效度样本的差距较大。

从这些数据可以看出,同一个测量模型(或测量工具)在两个分公司的施测上,复核效化的结果是获得支持的,甲分公司的测量结论可以套用到乙分公司。但是效度样本的测量拟合度则稍差,在套用结果时宜注意效度样本的特性。

第七节 结 语

本章内容介绍了多样本结构方程模型分析。从前述的介绍当中,我们可以发现多样本 SEM 分析可以说是基本的 SEM 分析的延伸。然而,不论原理概念或操作技术,多样本模型分析都具有相当高的价值,应用范围甚广,但是也具有一些限制。例如,数据库本身需要具有多重取样的程序,具有两个以上的样本数据。两个样本皆必须同时符合 SEM 分析的基本假设。当两个样本进行整合比较时,也必须符合共同比较的假设。

值得注意的是,即使多样本 SEM 可以处理外加的一个类别变量的影响,但是 SEM 分析的本质仍是在处理连续变量之间的共变结构。多样本 SEM 所增加的多样本比较,仅是将原本的 SEM 分析中的单一共变结构切割成多个平行共变结构,探讨共变结构在样本间的等同性。因此,多样本 SEM 不宜处理过多的样本,否则整个共变结构的分析会非常的复杂且耗费时间。

第十章

平均数结构分析

在结构方程模型分析中，主要的参数都是回归系数或近似于回归系数的参数，以及反映变量变异特性的方差与协方差。此时，变量分数的高低并不是分析的重点，各参数可以利用测量变量的方差协方差矩阵（S 矩阵）来加以估计，变量的分数高低（平均数）数据并不需要被 SEM 模型处理。但是，如果一旦研究者想要处理参数对于变量的平均数的影响时，传统上以 S 矩阵进行的 SEM 分析就不敷所需，而需使用本章所介绍的平均数结构分析。此外，如果研究者采用了重复量数设计或纵贯研究设计，关心多次测量的分数变动，也就是对于同一组受测者或受测物进行多次的测量，分析平均数的变化，也必须使用平均数结构分析。

第一节 平均数结构分析的原理

平均数结构分析可以说是多样本 SEM 的一个特色，类似于典型的 t 检验或方差分析，可以用来检验不同样本间的平均数差异。一般而言，多样本 SEM 在检验平均数结构时，系针对潜在变量的平均数进行检验，因为测量变量的存在是为了产生潜在变量，因此测量变量的程度差异是被忽略的。另外，由于平均数估计的主要元素是方程式的截距，而共变结构所估计的是斜率，因此平均数结构分析必须以共变结构分析为基础，延伸至平均数结构的讨论。再则，平均数结构必须在有多个样本情况下，才能产生多个平均数来比较，因此，平均数结构估计必须搭配多样本 SEM 分析。

由于多样本 SEM 检验在模型中纳入了一个类别变量，这个类别变量将整组的 SEM 矩阵切割成多个平行矩阵，以进行样本间的恒等性评估，但是多样本分析所关心的是共变结构所推导出来的参数在不同样本间的比较，并未涉及分数水平的比较。也就是说，类别变量将单一协方差矩阵切割成独立的协方差矩阵，观察不同样本下的共变结构的差异性，其本质仍为协方差，而样本间的差异也可能反映在模型中各变量的水平差异，例如潜在变量的得分高低，此时所分析的是不同样本的平均数结构（mean structure），而非传统 SEM 分析当中的共变结构（covariance structure）。

一、平均数结构的统计原理

平均数结构的 SEM 分析的特性,是在共变结构分析的基础之上,加入了平均数结构的估计,其模型构成的数学原理可由下列的通式来表示:

$$\eta = \alpha + B\eta + \Gamma\xi + \zeta \quad (10.1)$$

$$y = \tau_y + \Lambda_y\eta + \varepsilon \quad (10.2)$$

$$x = \tau_x + \Lambda_x\xi + \delta \quad (10.3)$$

上面各方程式与传统 SEM 模型方程式的差异在于增加了一个截距向量 (α 、 τ_x 、 τ_y 项),这些截距的资料反映了测量变量的分数水平的差异所造成的影响。其中最重要的是 α 向量。它反映了潜在变量的平均数高低。一般传统的共变结构 SEM 分析并不处理截距的估计,也就是将截距设定为 0 ($\tau_x=0$, $\tau_y=0$),因而造成潜在变量的截距向量为 0 矩阵 ($\alpha=0$)。

基于 SEM 的假设,带有平均数结构的结构模型方程式 10.1 中,解释残差 (ζ) 与内生潜在变量 (ξ) 为相互独立,带有平均数结构的测量模型方程式 (10.2) 与 (10.3) 中,测量残差 (ε 、 δ) 与外源潜在变量 (η 、 ξ) 亦相互独立。由于残差项为完全随机变量,因此期望值为 0,亦即 $E(\zeta)=0$ 、 $E(\varepsilon)=0$ 、 $E(\delta)=0$ 。但是潜在变量的分数并不是完全随机变量,潜在变量期望值 $E(\eta)$ 与 $E(\xi)$ 反映的即是平均数,平均数向量以 κ 表示。将公式 (10.1)、(10.2)、(10.3) 取期望值,得到结构模型与测量模型的平均数向量关系式如下:

$$E(\eta) = (I - B)^{-1}(\alpha + \Gamma\kappa) \quad (10.4)$$

$$E(y) = \mu_y = \tau_y + \Lambda_y(I - B)^{-1}(\alpha + \Gamma\kappa) \quad (10.5)$$

$$E(x) = \mu_x = \tau_x + \Lambda_x E(\xi) = \tau + \Lambda_x \kappa \quad (10.6)$$

在单一样本的 SEM 分析中,若 SEM 模型界定了截距来反映变量平均数的高低,对于基于共变结构所估计出来的各项参数并无影响。但是,在多样本 SEM 分析中,由于各参数除了反映斜率的差异之外,多了一个截距的比较,使得 SEM 分析的结果多了一个影响参数变异的来源。因此,截距的设定必须配合多样本分析进行才具有实质的意义。以外源变量的测量模型方程式为例,具有多样本的外源测量变量方程式可以写作:

$$x_g = \tau_x + \Lambda_{xg} \xi_g + \delta_g \quad (10.7)$$

在一个具有 g 个样本与 q 个测量变量的多样本 SEM 分析中, q 个测量变量形成了 k 个潜在变量,在加入了截距的设定后, τ_x 项反映了 q 个测量变量的截距向量组合。其数学模型如下:

$$E(x_g) = \tau + \Lambda_{xg} \kappa_g \quad (10.8)$$

其中 κ_g 就是代表第 g 个样本的各潜在变量的平均数,该项也是由具有 k 个元素的向量所组成。这些方程式的改变,虽然仅增加了截距项的估计,但是其估计过程与前几个章节的内容并无二致,但由于估计的参数增加,也将造成模型的复杂度增加,识别问题也就随之而来。

二、平均数结构分析的拟合函数

当多样本 SEM 模型当中加入了截距的设定,拟合函数 (fit function) 的计算必须反映样本平均数矩阵的影响,并考虑到不同样本之间的样本数的差异。特定组下拟合函数 F_g 通式如下:

$$F_g = (s_g - \sigma_g)' W_g^{-1} (s_g - \sigma_g) + (\bar{z}_g - \mu_g)' V_g^{-1} (\bar{z}_g - \mu_g) \quad (10.9)$$

其中 \bar{z}_g 为各样本平均数向量, μ_g 为各组的总体平均数, 可由公式 (10.5) 与 (10.6) 中, 各测量变量总体平均数 μ_{yg} 与 μ_{xg} 推导之, $\mu_g = (\mu_{yg}, \mu_{xg})'$ 。当四个截距与平均数矩阵被设定为 0 时, 公式 (10.9) 当中的第二项为 0, 公式 (10.9) 即还原为第九章的多样本 SEM 分析的拟合函数关系式。多样本综合拟合函数通式则为:

$$F = \sum_{g=1}^G \frac{N_g}{N} F_g \quad (10.10)$$

当使用 ULS、GLS、WLS 等参数估计程序时, 加权矩阵直接套入各组实际测量得到的协方差矩阵, $V_g = S_g$; 当使用 ML 估计程序时, 加权矩阵则使用各组估计协方差矩阵, $V_g = \hat{\Sigma}_g$ 。

第二节 平均数结构的分析技术

一、LISREL 的矩阵概念

从上述的方程式原理, 我们可以得知平均数结构的 SEM 分析新增四个参数矩阵 τ_x 、 τ_y 、 α 与 κ 矩阵, τ_x 、 τ_y 为测量模式的截距矩阵, α 为内生潜在变量的截距矩阵 (反映内生潜在变量的平均数), κ 则为外源潜在变量的平均数矩阵。在 LISREL 语法中, 这四个矩阵的默认值为 0。当用户欲进行平均数结构分析时, 仅需将这四个矩阵在 MModel 指令中适当地设定估计条件, 即可处理潜在变量的平均数估计。这四个矩阵的 LISREL 设定方式如表 10.1。

在 LISREL 语法中, 这四个矩阵的估计形式可为 FI (固定参数)、FR (自由估计参数)、IN (恒等参数)、PS (与前一组样本具有相同形态)、SS (与前一组样本具有相同起始值)、SP (与前一组样本具有相同形态与起始值)。

表 10.1 LISREL 平均数结构分析的矩阵概念列表

符号与发音	缩写	代表意义	<i>mm</i>	<i>mf</i>	order
结构模型矩阵					
α Alpha	AL	内生潜在变量的截距矩阵	ZE	FI	NE×1
κ Kappa	KA	外源潜在变量的平均数矩阵	ZE	FI	NK×1
τ_y Tau-y	TY	内生测量变量的截距矩阵	ZE	FI	NY×1
τ_x Tau-x	TX	外源测量变量的截距矩阵	ZE	FI	NX×1

二、平均数结构分析的识别性

由于平均数结构是在原有的 SEM 模型中外加的条件，因此在计算识别性问题时，平均数结构本身的设定条件需仔细的评估，而且因为平均数结构的加入，导致新增的估计参数也可能造成模型识别不足的问题。因此，在执行多样本平均数结构设定时，必须仔细评估模型的设定条件与识别性是否恰当。

在识别性的计算上，一个比较明显的问题是样本间因素恒等性会造成平均数结构无法识别的问题（Sörbom, 1974）。由于潜在变量的平均数是由测量变量的分数所计算得出，因此 g 个样本的测量模型的因素恒等性是相当可能存在的。一旦因素恒等性存在时，公式（10.5）与（10.6）的期望式即无法成立。Sörbom（1974）以一个简单的方程式来说明因素恒等性假设的影响：

$$E(x_g) = \tau_x - \Lambda_x d + \Lambda_x (\kappa_g + d) = \tau_x + \Lambda_{xg} \kappa_g \tag{10.11}$$

上式是在公式(10.2)中加入了潜在变量的平均数向量 κ_g 中再加入一个 k 元素的 d 向量,然后在截距中扣回 d 向量加入的影响,公式(10.3)将仍可以产生公式(10.2)的导出式,因此,公式(10.2)若要能够被识别,必须在式中增加一组 k 个设限,使平均数结构得以定位(或作为他组样本的参照)。一般的做法是将其中一个样本的 κ_g 向量设定为 0 ($\kappa_g=0$),使其他的各组的平均数得以自由估计,整个模型方才获得成功的识别。

Jöreskog 与 Goldberger (1975)提出了一个 MIMIC 模型 (multi-indicators multi-causes model) 进行多重指标与多重因果模型的跨样本分析。MIMIC 模型适用于两个样本的 SEM 比较,其做法是将双样本的切割设定为一个二分虚拟变量(例如 0 代表第一个样本,1 代表第二个样本),纳入 SEM 模型作为一个独立变量,使原有的 SEM 模型中增加了一个外源测量变量,而输入的协方差矩阵也因此增加了一组二分虚拟变量的协方差与方差,但仅需进行单样本共变结构分析,组间差异反映在新增的二分虚拟变量的参数估计中。

以 MIMIC 模型处理双样本分析,虽然会增加参数估计数目,但是由于增加一个观察变量,使得共变结构的数据点数也同时增加,因此不会造成识别性不足的问题。而由于虚拟变量的纳入,使得双样本数据可以改为单样本模式分析,因此操作上较前述的模式简易。此外, MIMIC 模型的优点是可以延伸到将其他类型的虚拟变量纳入 SEM 分析,例如应用到实验设计中的不同组别方差分析检验。

第三节 平均数结构分析: 测量模型

平均数结构分析属于多样本 SEM 分析的一种应用。其基本原理是在两个以上的多样本 SEM 分析中,当因素恒等性存在的情况下,以其中一组为基准,输入各测量变量的平均数数据后,藉以估计其他组的潜在变量的平均数高低,并可利用显著性检验来检验不同组间的潜在变量平均数是否具有显著差异。

在先前的范例中,我们检验了《组织创新气氛量表》十八题短题本的因素结构在性别变量上的恒等性。结果指出因素恒等性并不存在,但是为了示范潜在变量的平均数估计,本范例则延续前一个范例的数据,将不同性别之间的验证性因素分析,再增加平均数结构的估计来示范如何进行平均数结构的估计。但是在正常的情况下,如果因素的恒

等性并不存在时,潜在变量的平均数估计不宜进行。换句话说,平均数结构分析必须建立在因素恒等性的假设之上,平均数的差异才具有统计上的比较价值。

根据前一章的说明,因素恒等性的存在有多种不同的层次,当用于潜在变量的平均数估计时,跨样本之间的因素结构在因素载荷与残差两个矩阵需具有恒等性,也就是两个样本的 Λ 矩阵与 Θ 矩阵被假设为相等同(Jöreskog & Sörbom, 1996)。

一、模型界定

本范例所进行的平均数结构分析,在因素模型上仍为一个以 18 个测量变量来估计六个因素的 CFA 模型,所不同的是增加对于跨样本的测量模型中进行截距项的估计,并推导出代表潜在变量的平均数的 Kappa 矩阵的估计数。

以 18 个测量变量所建立的 CFA 模型可以产生 171 个测量数据点数(DP=171),由于包含多组 SEM 模型,因此总 DP 值为各单一模型的 k 倍(k 为样本数),本范例有男女两个样本,DP 值为 342。

在单一样本的独立 CFA 模型中,18 个测量变量来测量六个因素的 CFA 模型共有 51 个被估计的参数,两个样本总估计参数为 102。进行平均数结构估计时需增加两项设定:

1. 观察变量的截距被估计,共有 18 个 Tau 参数需被估计。
2. 潜在变量的平均数被估计,共计有 6 个 Kappa 参数需被估计。

模型中由于因素载荷(18 个)与残差方差(18 个)被设定为恒等,因此共有 36 个参数被固定为 0 不予估计,无平均数设定下的参数数目为 66,自由度为 $342-66=276$,作为基准模型。增加平均数估计设定的跨样本 CFA 模型,则需增加 18 个截距与 6 个平均数。

二、分析语法与参数估计

表 10.2 列出了测量模型的平均数结构分析的 LISREL 与 SIMPLIS 语法,其中 LISREL 语法的第 1 至第 9 行用以设定男性样本模型,第 13 至第 20 行则用以设定女性样本模型,两个样本的因素结构完全相同。

表 10.2 多样本平均数结构分析语法

LISREL 语法范例(Ch10a.ls8)

```

1  GROUP: MALE
2  DATA  NG=2 NI=18 NO=172
3  KM SY FI=Ch10am.cor; ME FI=Ch10am.cor; SD FI=Ch10am.cor
4  LA; A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3
5  MODEL NX=18 NK=6 PH=SY,FR TD=DI,FR LX=FU,FI TX=FR KA=FI
6  LK; VALUE JOBSTYLE TEAMWORK LEADERSHIP LEARNING ENVIRONMENT
7  FREE LX(1,1) LX(2,1) LX(3,1) LX(4,2) LX(5,2) LX(6,2) LX(7,3) LX(8,3)
   LX(9,3)
8  FREE LX(10,4) LX(11,4) LX(12,4) LX(13,5) LX(14,5) LX(15,5)
   LX(16,6)LX(17,6) LX(18,6)
9  VALUE 1 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
10 PD
11 OUTPUT SS SC
12 GROUP: FEMALE
13 DATA  NO=140
14 KM SY FI=Ch10af.cor; ME FI=Ch10af.cor; SD FI=Ch10af.cor
15 LA; A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3
16 MODEL LX=IN TD=IN TX=IN KA=FR
17 LK; VALUE JOBSTYLE TEAMWORK LEADERSHIP LEARNING ENVIRONMENT
18 FREE LX(1,1) LX(2,1) LX(3,1) LX(4,2) LX(5,2) LX(6,2) LX(7,3) LX(8,3)
   LX(9,3)
19 FREE LX(10,4) LX(11,4) LX(12,4) LX(13,5) LX(14,5) LX(15,5)
   LX(16,6)LX(17,6) LX(18,6)
20 VALUE 1 PH(1,1) PH(2,2) PH(3,3) PH(4,4) PH(5,5) PH(6,6)
21 PD
22 OUTPUT SS SC
23 End of Problem

```

SIMPLIS 语法范例(Ch10a.spl)

```

1  GROUP: MALE
2  Observed variables: A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2
   F3
3  Raw data from file Ch10am.dat
4  Sample Size = 172
5  Latent Variables  VALUE JOBSTYLE TEAMWORK LEADERSH LEARNING ENVIRONM
6  Relationships
7  A1-A3 = CONST VALUE
8  B1-B3 = CONST JOBSTYLE
9  C1-C3 = CONST TEAMWORK

```

```

10  D1-D3 = CONST LEADERSH
11  E1-E3 = CONST LEARNING
12  F1-F3 = CONST ENVIRONM
13  Set the Variance of VALUE-ENVIRONM to 1.00
14  Path Diagram
15  GROUP: FEMALE
16  Observed variables: A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2
    F3
17  Raw data from file Ch10af.dat
18  Sample Size = 140
19  Latent Variables  VALUE JOBSTYLE TEAMWORK LEADERSH LEARNING ENVIRONM
20  Relationships
21  VALUE = CONST
22  JOBSTYLE = CONST
23  TEAMWORK = CONST
24  LEADERSH = CONST
25  LEARNING = CONST
26  ENVIRONM = CONST
27  Set the Variance of VALUE-ENVIRONM to 1.00
28  Set the Covariance of JOBSTYLE and VALUE Free
29  Set the Covariance of TEAMWORK and VALUE Free
    ..... (略)
30  End of Problem

```

LISREL 语法所输入的数据为相关系数矩阵，男性样本的相关矩阵数据存放 Ch10am.cor 档案中，女性样本的相关矩阵数据存放 Ch10af.cor 档案中。值得注意的是，在这两个数据文件中的最后两行是平均数与标准差的数据。在估计平均数结构时，必须让计算机读入观察变量的平均数数据，否则无法进行截距的估计。

恒等性设定位于第 16 行的 LX=IN 与 TD=IN，表示因素载荷与残差具有跨样本的恒等性。平均数结构的设定位于第 5 行与第 16 行。其中第 5 行的 TX=FR 表示在第一个样本（男性样本）中，测量变量的截距进行自由估计，KA=FI 表示因素的平均数被限定为 0；第 16 行的 TX=IN 指出截距项具有跨样本的等同性，KA=FR 表示女性样本的潜在变量的平均数进行估计。在 SIMPLIS 语法部分，所读入的数据为原始数据，分别是 Ch10am.dat 与 Ch10af.dat。对于平均数结构的设定，在测量模型位于第 7 至第 12 行，各行均增加了 CONST，表示观察变量的截距纳入估计。结构模型的潜在平均数设定位于第 21 至第 26 行，以第 21 行为例，VALUE=CONST 表示潜在变量的截距纳入估计。

在男性样本中没有潜在变量截距的设定，表示以男性样本的潜在平均数设定为

0，作为参照组。模型中设定跨样本等同的则为因素载荷，在 SIMPLIS 的女性模型部分没有出现因素载荷参数，表示设定为跨样本相等。

三、估计结果与分析

本范例的特点在于增加了截距的估计，因此 LISREL 将读入测量变量的平均数，以作为截距估计的基础。在报表上，我们可以看到两个样本的读入数据，在协方差矩阵之后会列出测量变量的平均数数据。以下仅列出男性样本的报表（女性样本的报表形式类似）。

（一）读入资料检查

Covariance Matrix						
	A1	A2	A3	B1	B2	B3
A1	0.83					
A2	0.42	0.92				
A3	0.34	0.28	0.81			
.....(略)						
Means						
	A1	A2	A3	B1	B2	B3
	4.55	4.44	4.15	4.11	4.41	4.39
Means						
	C1	C2	C3	D1	D2	D3
	4.44	4.33	4.44	4.83	4.95	4.91
Means						
	E1	E2	E3	F1	F2	F3
	4.75	4.88	4.80	4.30	4.81	4.70

在模型界定报表中，则列出了 18 个被估计的截距参数（编号 52 至 69），以及 6 个被估计的平均数参数（编号 85 至 90），如下所示：

TAU-X

A1	A2	A3	B1	B2	B3
52	53	54	55	56	57

TAU-X

C1	C2	C3	D1	D2	D3
58	59	60	61	62	63

TAU-X

E1	E2	E3	F1	F2	F3
64	65	66	67	68	69

KAPPA

VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
85	86	87	88	89	90

(二) 参数估计结果

平均数结构的 CFA 分析最重要的参数是截距 (Tau 参数) 与潜在变量平均数 (Kappa 参数的估计)。利用这些参数估计数的标准误, 则可以计算出显著性检验的统计量 (t 值) (注 1)。

TAU-X

A1	A2	A3	B1	B2	B3
4.55 (0.07) 62.71	4.43 (0.07) 61.73	4.17 (0.07) 63.47	4.14 (0.08) 52.07	4.41 (0.08) 52.27	4.38 (0.08) 56.06

TAU-X

C1	C2	C3	D1	D2	D3
4.43 (0.07) 63.24	4.40 (0.07) 59.86	4.36 (0.07) 60.48	4.85 (0.07) 69.57	4.96 (0.06) 79.16	4.84 (0.06) 76.53

注 1: 截距参数的估计数、标准误与显著性统计量, 大于 1.96 表示截距显著不等于 0。估计数的数值应近似于测量变量的平均数。

TAU-X

E1	E2	E3	F1	F2	F3
4.75 (0.07) 67.77	4.87 (0.07) 65.00	4.82 (0.07) 69.26	4.38 (0.08) 58.09	4.82 (0.08) 62.63	4.68 (0.07) 68.22

KAPPA

VALUE	JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
-0.38 (0.13) -2.90	-0.37 (0.13) -2.93	-0.19 (0.13) -1.43	-0.04 (0.12) -0.31	-0.35 (0.12) -2.84	-0.56 (0.13) -4.27

注 2：潜在变量的平均数估计数，以男性样本（潜在变量平均数为 0）作为基准值，所计算出的跨样本差异数。负值表示女低于男，t 值小于-1.96 表示具有统计的显著性。

值得注意的是，Kappa 系数为女性样本潜在变量平均数的估计数，男性样本的 Kappa 系数被设定为 0，以作为与女性样本平均数的参照。当数值为正值时，表示女性样本的平均数高于男性，当数值为负值时，表示女性样本的平均数低于男性，t 值若高于±1.96，表示平均数的差异达到.05 的统计水平。数据显示，除了团队合作与领导风格两因素外，另外四个因素平均数在与男性样本的参照下，具有统计的意义，也就是女性样本的因素平均数显著低于男性（注 2）。

（三）模型拟合度分析

由于平均数结构的检验是基于因素恒等性存在的情况下来进行测量变量截距的估计与潜在变量平均数的估计，因此，基准模型应为无平均数结构设定的恒等模型，替代模型为以表 10.2 之 LISREL 语法所定义的平均数结构 CFA 模型。由于基准模型已于先前估计过，因此直接取用先前的数据，而平均数结构模式的模型拟合度数据则经过上述 LISREL 语法估计后可以得到，经过整理后列举于表 10.3。

由表 10.3 的数据显示，两个 CFA 模型的拟合指数都十分接近，而两者的卡方值的差异量未达显著水平 ($\Delta\chi^2=14.49$, $df=12$, $p>.05$)，表示增加平均数与截距参数并不影响模型拟合度的计算。但是值得注意的是，由于增加截距与平均数参数并未改变 CFA 模型的结构，仅增加自由参数数目，因此卡方差异量单纯反映了 SEM 技术

复杂度提高后的影响, 而和假设模型与实际观察数据的拟合无关。

表 10.3 平均数结构 CFA 分析模型拟合度评估摘要表

Model	χ^2	df	RMSEA	NNFI	CFI
基准模型	463.39 ($p=.000$)	276	.066	.96	.96
平均数模型	477.88 ($p=.000$)	288	.065	.96	.96
差异检验	$\Delta=14.49$ ($p>.05$)	$\Delta=12$			

四、参数整理与报告

由前述的数据可以看出, 平均数结构的 CFA 分析并不影响 CFA 模型本身的各参数 (例如因素载荷、残差变异量与相关系数) 的估计。因此, 在报告最后的结论时, 仅需列出截距与平均数估计的结果。

表 10.4 跨样本平均数结构 CFA 分析的参数估计结果

因素	题目	因素载荷 λ	残差 δ	截距 τ	因素平均数 κ	显著性检验 (t)
组织 价值	A1	.82	.33	4.55	-.38	-2.90**
	A2	.70	.51	4.43		
	A3	.65	.60	4.17		
工作 方式	B1	.68	.54	4.14	-.37	-2.93**
	B2	.82	.33	4.41		
	B3	.79	.37	4.38		
团队 合作	C1	.72	.48	4.43	-.19	-1.43
	C2	.72	.48	4.40		
	C3	.65	.58	4.36		
领导 风格	D1	.87	.24	4.85	-.04	-0.31
	D2	.88	.22	4.96		
	D3	.72	.48	4.84		
学习 成长	E1	.83	.31	4.75	-.35	-2.84**
	E2	.90	.19	4.87		
	E3	.80	.35	4.82		

(续表)

因素	题目	因素载荷 λ			残差 δ	截距 τ	因素平均数 κ			显著性检验 (t)			
环境 气氛	F1	.52			.73	4.38	-.56			-4.27**			
	F2	.74			.45	4.82							
	F3	.84			.30	4.68							
因素相关		男性样本						女性样本					
组织价值		1.00						1.00					
工作方式		.63	1.00					.44	1.00				
团队合作		.63	.64	1.00				.36	.76	1.00			
领导风格		.64	.43	.48	1.00			.23	.49	.58	1.00		
学习成长		.58	.53	.63	.53	1.00		.46	.61	.57	.61	1.00	
环境气氛		.75	.45	.66	.30	.39	1.00	.62	.27	.52	.35	.44	1.00

* $p < .05$ ** $p < .01$

在列举这些数据时,应采用跨样本的完全标准化解,也就是 Common metric completely standardized solution,才能符合两样本具有跨样本等同性的基本假设的精神。也因为跨样本之间在因素载荷与残差具有等同,因此两个样本的终解在 LAMBDA-X 与 THETA-DELTA 参数均为等同,两个样本的因素相关是独立估计(注3与注4),在整理最后的结论时,应分别列举两个样本的因素相关。所有的参数估计结果与显著性检验结果,可以整理成表 10.4 形式的摘要表,以利阅读。

Common Metric Completely Standardized Solution

LAMBDA-X		JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
VALUE						
A1	0.82	---	---	---	---	---
A2	0.70	---	---	---	---	---
A3	0.63	---	---	---	---	---
B1	---	0.68	---	---	---	---
B2	---	0.82	---	---	---	---
B3	---	0.79	---	---	---	---
.....(略)						
PHI		JOBSTYLE	TEAMWORK	LEADERSH	LEARNING	ENVIRONM
VALUE						
组织价值	1.00					
工作方式	0.63	1.00				
团队合作	0.63	0.64	1.00			
领导风格	0.64	0.43	0.48	1.00		
学习成长	0.58	0.53	0.63	0.53	1.00	
环境气氛	0.75	0.45	0.66	0.30	0.39	1.00

注3: 男性样本的因素相关。在跨样本间并未等同。

THETA-DELTA

A1

A2

A3

B1

B2

B3

0.33

0.51

0.60

0.54

0.33

0.37

.....(略)

GROUP: FEMALE

Common Metric Completely Standardized Solution

LAMBDA-X

VALUE

JOBSTYLE

TEAMWORK

LEADERSH

LEARNING

ENVIRONM

A1

A2

A3

B1

B2

B3

0.82

- -

- -

- -

- -

- -

0.70

- -

- -

- -

- -

- -

0.63

- -

- -

- -

- -

- -

- -

0.68

- -

- -

- -

- -

- -

0.82

- -

- -

- -

- -

- -

0.79

- -

- -

- -

- -

.....(略)

注 4：女性样本的因素相关。

PHI

VALUE

JOESTYLE

TEAMWORK

LEADERSH

LEARNING

ENVIRONM

组织价值

工作方式

团队合作

领导风格

学习成长

环境气氛

1.00

0.44

0.36

0.23

0.46

0.62

1.00

0.76

0.49

0.61

0.27

1.00

0.58

0.57

0.52

1.00

0.61

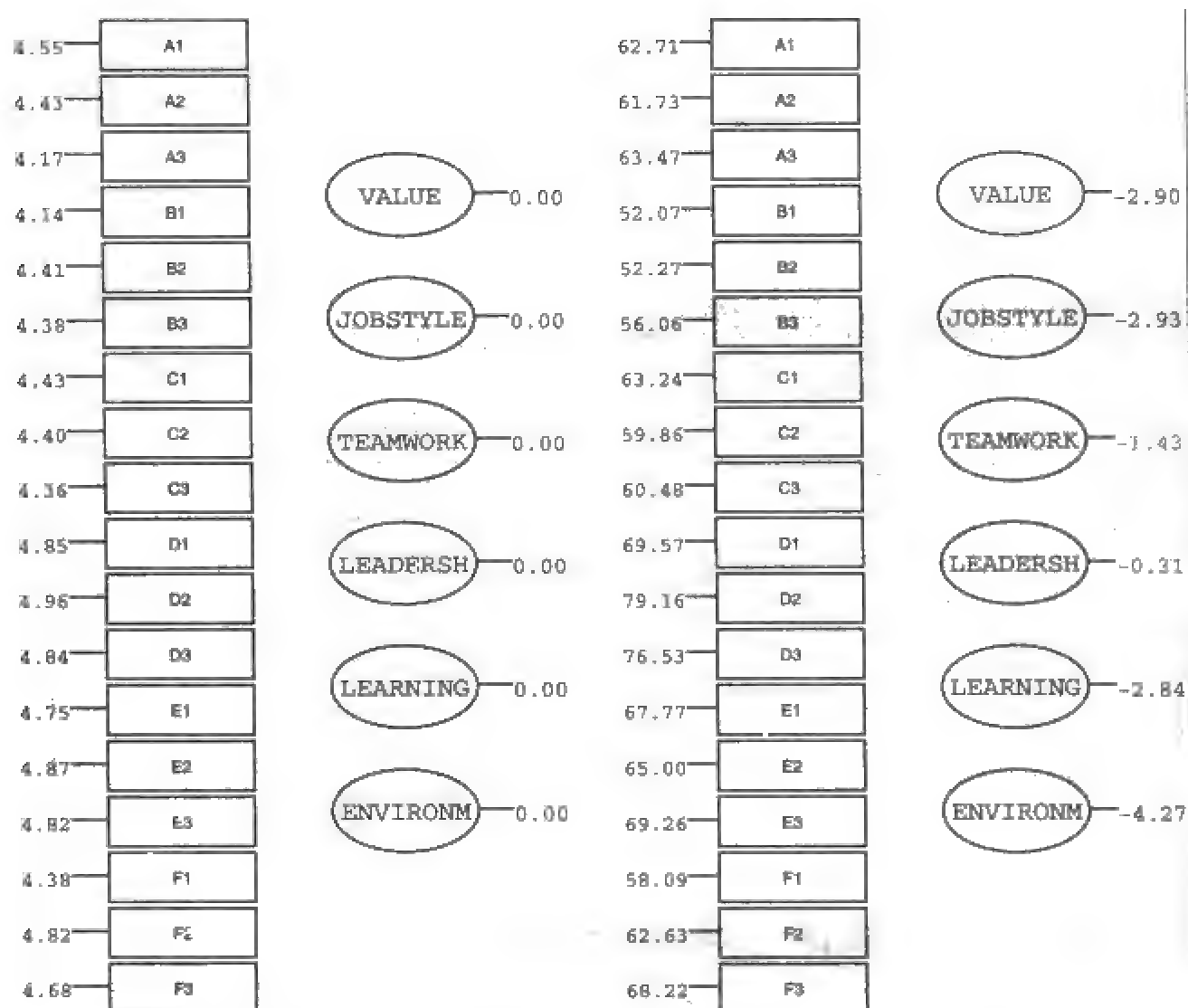
0.35

1.00

0.44

1.00

表 10.4 的数据清楚说明了六个因素平均数的性别差异，女性的因素得分均较男性为低，且在组织价值、工作方式、学习成长与环境气氛四个因素达到统计的显著性。相关系数的数据则显示男性与女性样本的因素相关明显不同。至于数据的详细说明不在此赘述。另外，也可以 LISREL 的路径图功能将各变量的平均数估计数以图 10.1 的方式呈现。在此可以清楚看到各变量平均数估计结果（图 10.1 所呈现为女性样本数据，男性样本的各平均数估计数因其作为参照之故为 0）。

图 10.1 跨样本平均数结构 CFA 分析各变量平均数与 t 检验值

五、AMOS 平均数结构的示范

延续前一章的 AMOS 多样本分析,本章仅是增加了截距与潜在变量平均数的估计, AMOS 的操作仅需增加一些设定,说明如下。

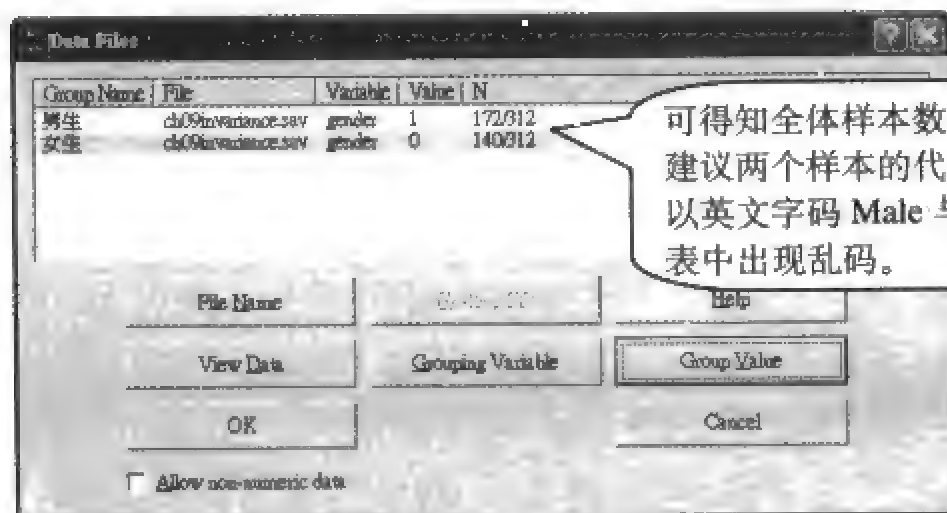
(一) 操作步骤

步骤一: 绘制假设模型图

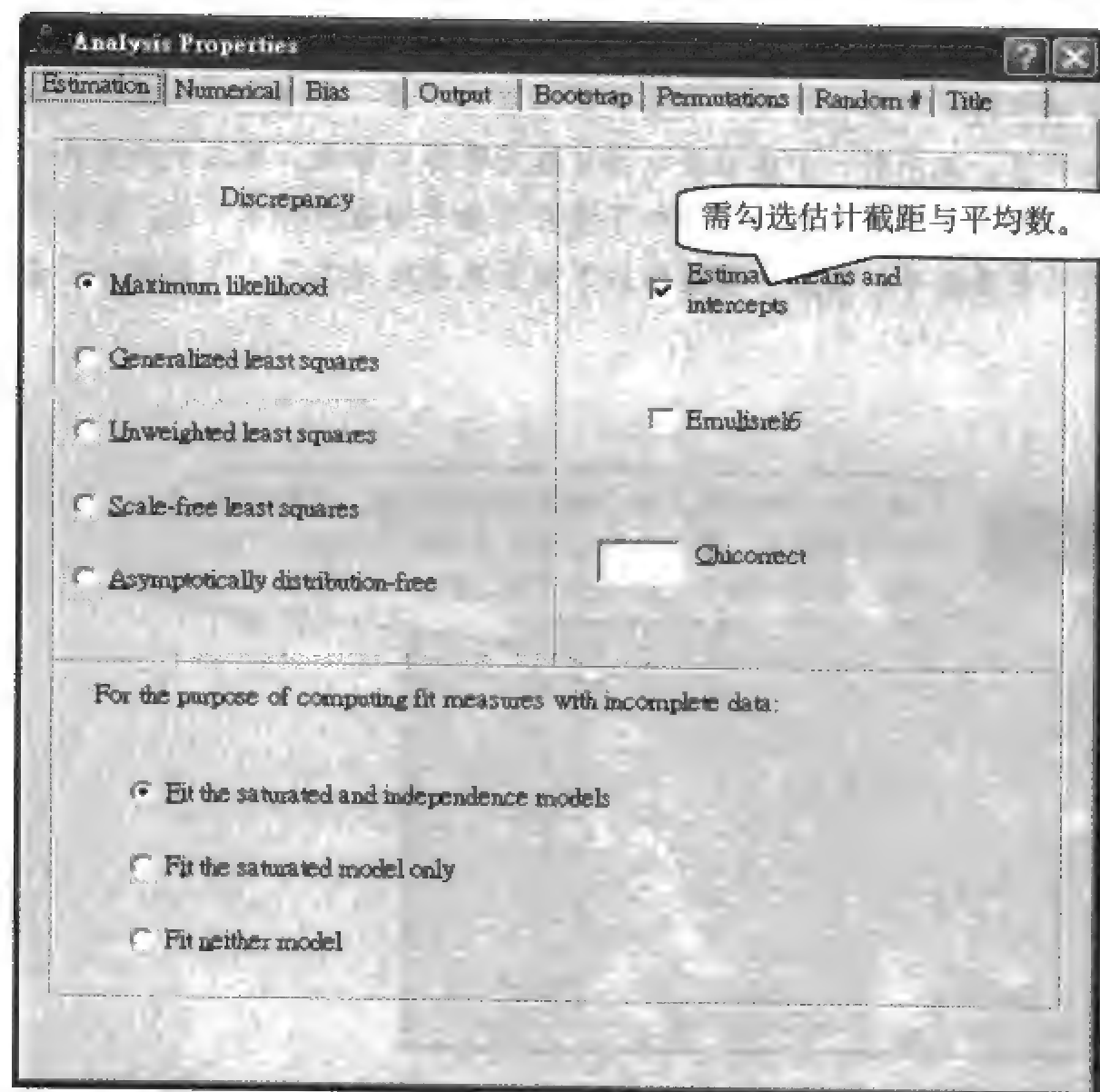
步骤二: 点选 Analyze->Manage Groups

步骤三：设定样本名称：男生与女生

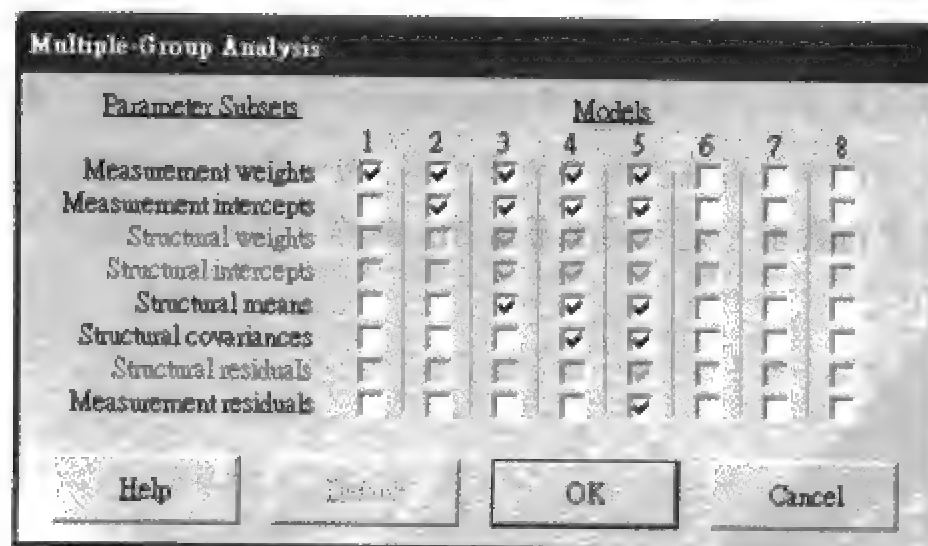
步骤四：点选数据文件，开启档案，设定两个样本的来源



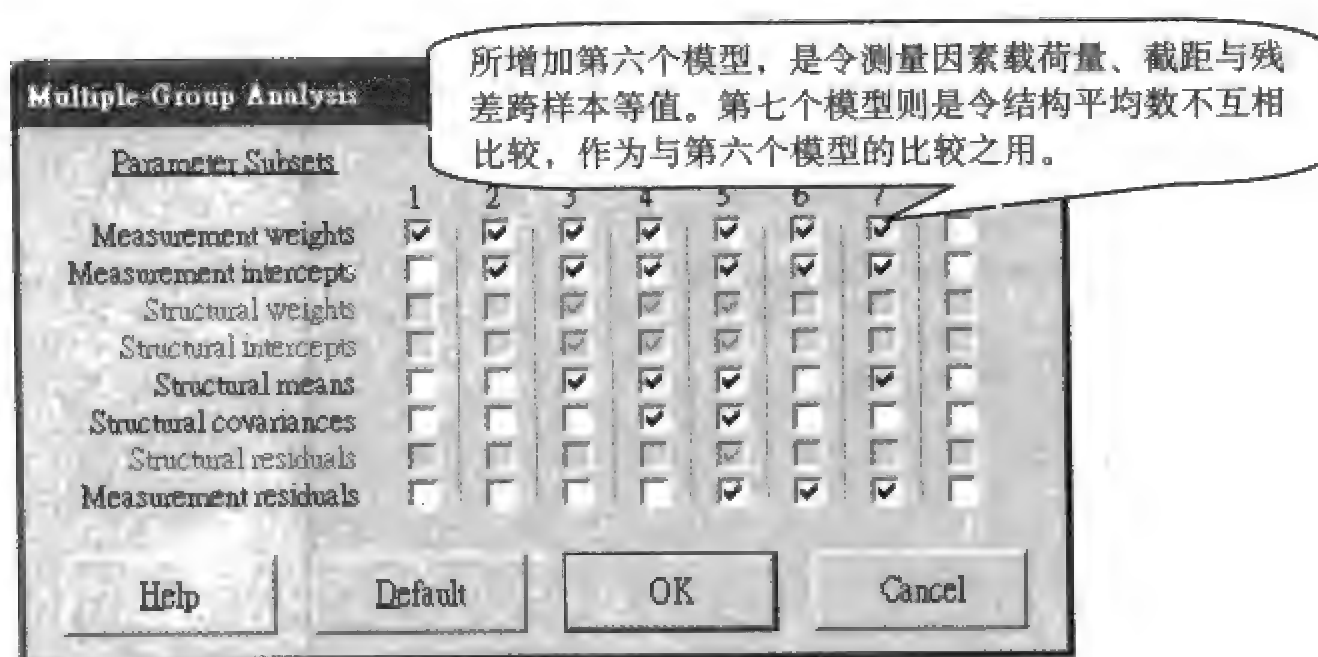
步骤五：点选分析选项，点选 Estimate means and intercepts，如此即宣告要进行截距与平均数的估计。



步骤六：点选  (多样本分析)，按确定之后即进入模型界定对话框如下：

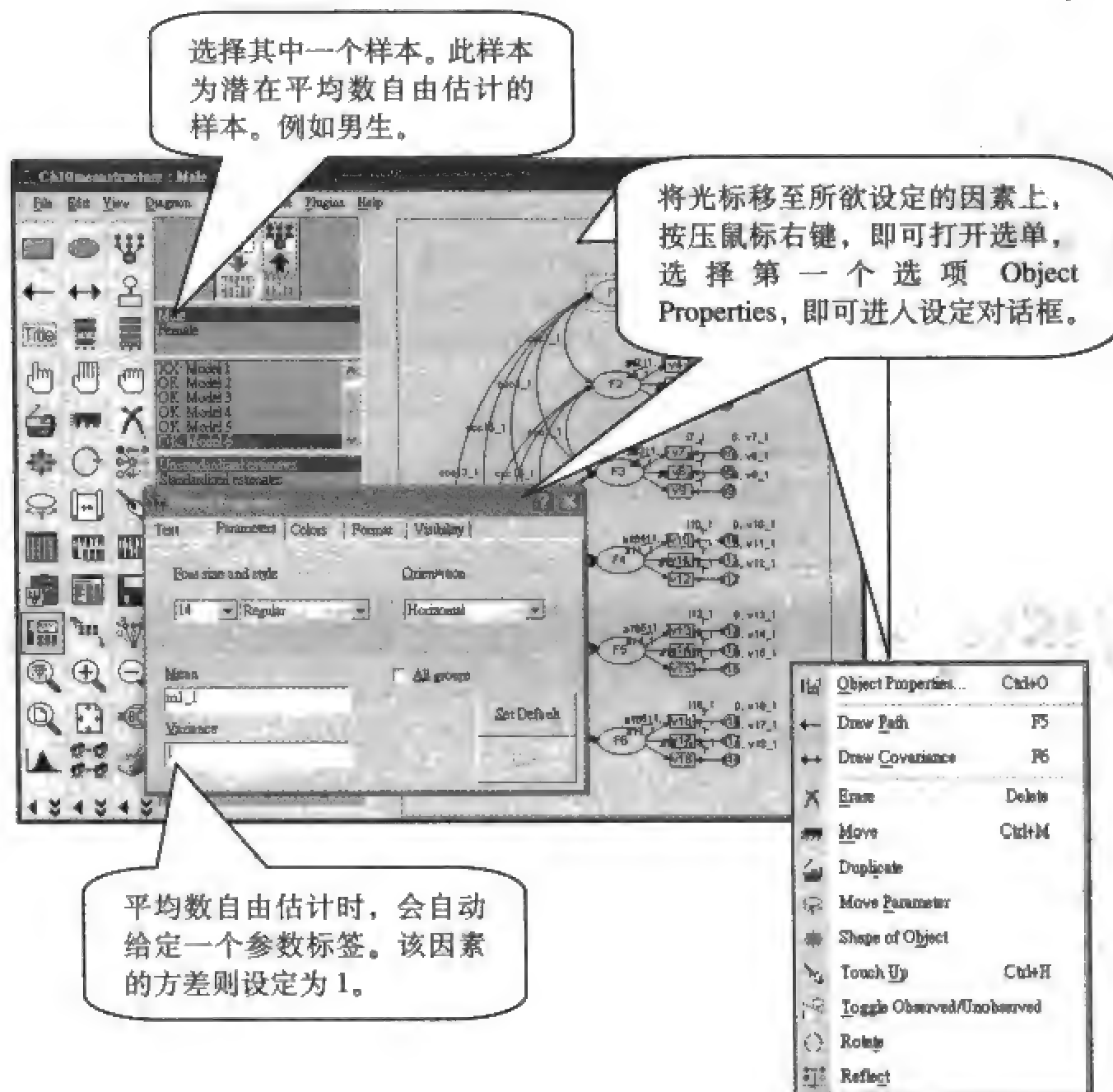


对话框中显示包含截距与潜在变量平均数的恒等模型。第一个模型表示两个样本的因素载荷跨样本等值。其中第一个模型表示因素载荷跨样本等值。第二个模型表示增加了观察变量截距恒等。第三个模型则增加了潜在变量平均数跨样本等值。第四个模型增加了因素的共变量恒等，第五个模型增加了测量残差跨样本等值，此即为完全等值模型。



如果想要在 Model 2 的基础之上先增加一个测量残差跨样本的限制（亦即本范例在 LISREL 范例中设定），即为下表的 Model 6 所示。第七个模型则是令结构平均数不互相比，作为与第六个模型的比较之用。

步骤七：设定其中一组结构平均数为 0。此步骤在平均数结构分析中非常重要。操作方法是选定两个样本中的其中一个样本的潜在变量平均数为零，另一个样本则自由估计，如此即可用来检验两个样本的平均数差异是否具有显著性。操作方式如下（每一个潜在变量都要重复此一操作）：



(二) 估计结果：模型拟合比较

进行估计后，AMOS 会将六个模型（包含未设限模型、饱和模型与独立模型）的估计结果列于报表中。下图即是 AMOS 分析的模式拟合指数：

Model6 是比照前面 LISREL 范例的设定（观察变量载荷、截距与残差为跨样本等值）， $\chi^2=497.372$ ， $df=288$ ， $p<.001$ ，结果与先前的 LISREL 估计结果相似。Model7 则是将 Model6 当中的结构平均数不加以比较后的模型（因此模式自由度增加了 6）， $\chi^2=525.369$ ， $df=294$ ， $p<.001$ ，模型拟合情形变得更不理想，表示结构的平均数应该加以估计。至于两个模型的嵌套比较（卡方差异检验）列于下。



Amos Output

chl Omeststructure.amw

- * Analysis Summary
- Notes for Group
- * Variable Summary
- Parameter summary
- * Notes for Model
- * Estimates
- Minimization History
- * **Model Fit**
- * Model Comparison
- Execution Time

Model

Female

Unconstrained model

Model 1

Model 2

Model 3

Model 4

Model 5

Model 6

Model 7

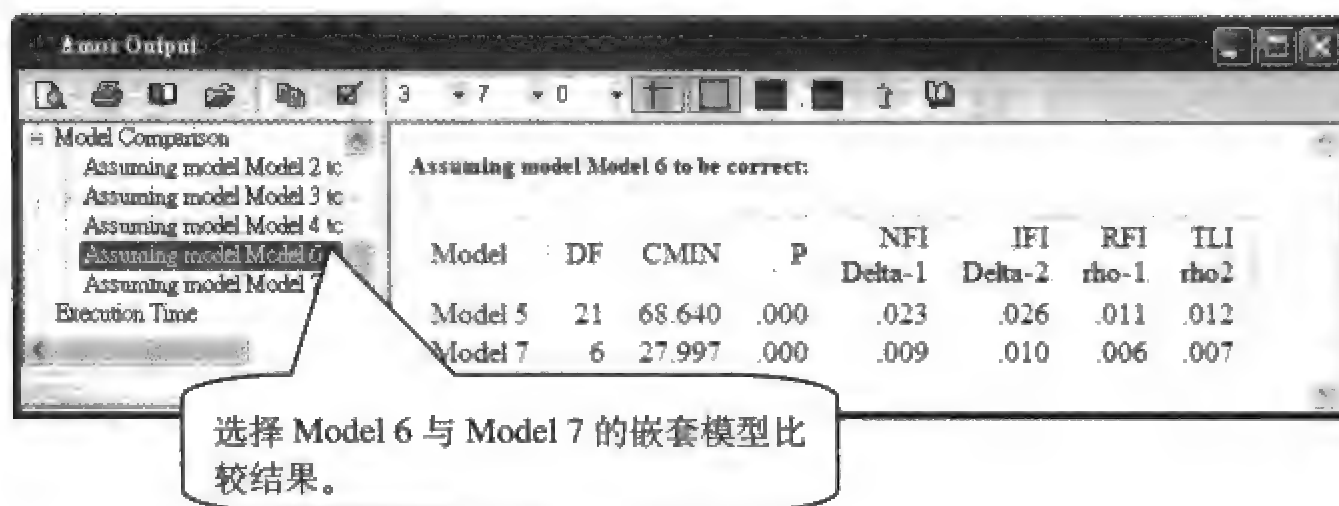
Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Model 2	108	465.399	270	.000	1.724
Model 3	102	493.050	276	.000	1.786
Model 4	87	536.486	291	.000	1.844
Model 5	69	566.012	309	.000	1.832
Model 6	90	497.372	288	.000	1.727
Model 7	84	525.369	294	.000	1.787
Saturated model	378	.000	0		
Independence model	72	2978.124	306	.000	9.732

Baseline Comparisons

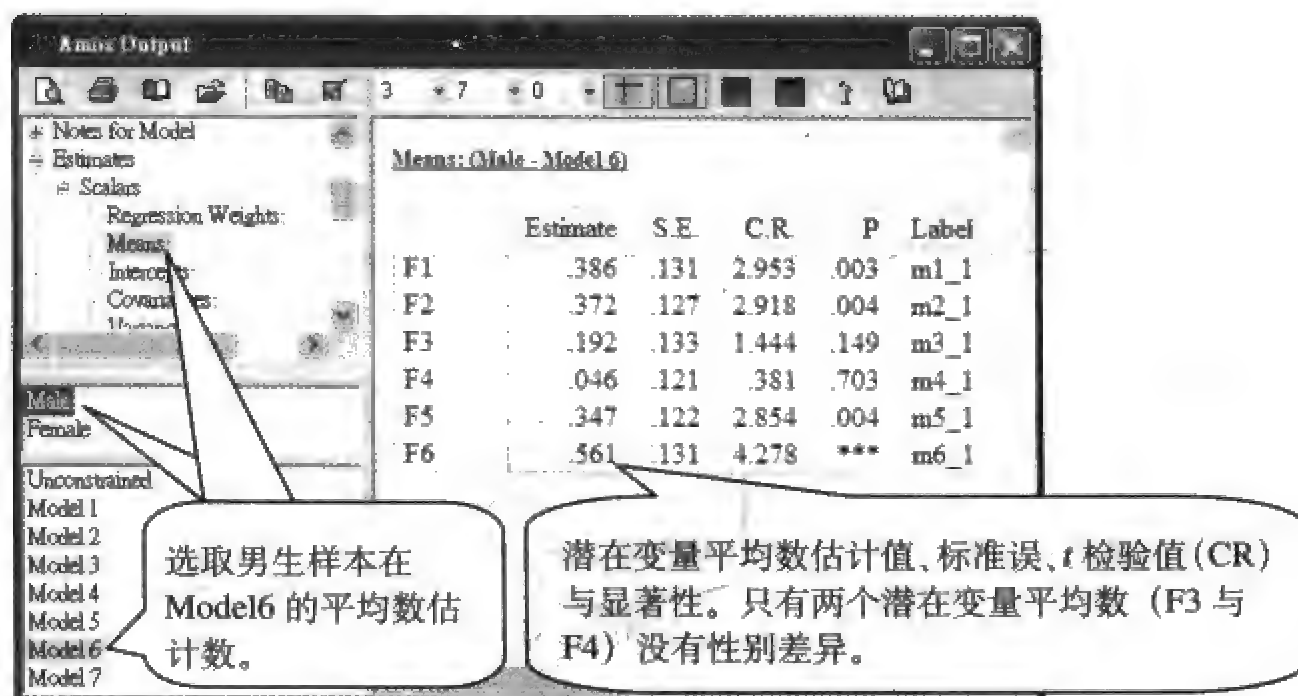
Model	NFI	RFI	IFI	TLI	CFI
	Delta1	rho1	Delta2	rho2	
Model 2	.844	.823	.928	.917	.927
Model 3	.834	.816	.920	.910	.919
Model 4	.820	.811	.909	.903	.908
Model 5	.810	.812	.904	.905	.904
Model 6	.833	.823	.922	.917	.922
Model 7	.824	.816	.914	.910	.913
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000



Model 6 与 Model 7 两个模型的卡方差异量 $\Delta\chi^2=27.997$, $\Delta df=6$, $p<.001$ 。表示增加结构平均数的比较能够有效改善模型拟合。

(三) 估计结果：平均数差异检验

平均数差异的检验主要并非检视模型拟合，而是比较两个样本在结构平均数上的差异是否具有统计意义。检视估计数当中的 Means 即可在结构平均数设定为自由估计样本（男生）得到下列报表。所得到的结果与先前 LISREL 估计结果相近。



六、Mplus 估计语法与结果

以 Mplus 来执行平均数结构分析的语法，本范例是在 Λ_x 、 Φ 、 Θ_s 三个矩阵均被限定为样本间相等时比较潜在变量的平均数差异。

(一) Mplus 语法

TITLE: Ch10a Mean structure modeling for Measurement model
DATA:
 NGROUPS=2;
 FILE (male) IS Ch10am.dat;
 FILE (female) IS Ch10af.dat;
VARIABLE:
 NAMES ARE a1-a3 b1-b3 c1-c3 d1-d3 e1-e3 f1-f3;
MODEL:
 VALUE by a1*a2-a3;
 JOBSTYLE by b1*b2-b3;
 TEAMWORK by c1*c2-c3;
 LEADERSH by d1*d2-d3;
 LEARNING by e1*e2-e3;
 ENVIRONM by f1*f2-f3;
 VALUE@1 JOBSTYLE@1 TEAMWORK@1 LEADERSH@1 LEARNING@1 ENVIRONM@1;
 a1(1); a2(2); a3(3); b1(4); b2(5); b3(6); c1(7); c2(8); c3(9); d1(10);
 d2(11); d3(12); e1(13); e2(14); e3(15); f1(16); f2(17); f3(18);
OUTPUT:
 STANDARDIZED;

多样本分析的设定格式
NGGROUP=2 表示有两个样本，数据各自存放于两个档案中。

模式设定
每一个潜在变量的第一个因素载荷量也自由估计，因此需将潜在变量方差设为 1。这些参数在两个样本都是等值设定。

观察变量截距恒等设定
两个样本的每一个观察变量截距设为相等，必须在每一个截距上设一个数字，放在括号内。

(二) Mplus 报表

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

模型拟合数据
卡方值为 500.581，与 LISREL 的 MIN 卡方值相近。

Value	500.581
Degrees of Freedom	288
P-Value	0.0000

Means				
VALUE	-0.386	0.130	-2.961	0.003
JOBSTYLE	-0.372	0.127	-2.926	0.003
TEAMWORK	-0.192	0.133	-1.437	0.151
LEADERSH	-0.046	0.121	-0.382	0.703
LEARNING	-0.347	0.121	-2.862	0.004
ENVIRONM	-0.561	0.131	-4.288	0.000

潜在变量的截距在第二个样本设定为自由估计。ENVIRONM 的估计值为-.561, t 值为-4.288, 达统计显著水平, 与 LISREL 估计结果相近。

第四节 平均数结构分析：统合模型

除了测量模型中得以进行潜在变量的平均数结构, 在结构模型中亦可以进行潜在变量的平均数结构的分析。不论是 CFA 或统合模型分析, 当因素的平均数作为研究的焦点之一时, 因素恒等性都是一个很重要的基础。由于统合模型不仅涉及测量模型的设定, 也具有结构模型, 因此, 统合模型的平均数结构分析会较 CFA 模型复杂。因为平均数结构设定不仅要进行跨样本的比较, 也涉及不同模型间恒等性的检验, 当恒等性的前提适当地满足时, 平均数的比较才有意义。

本范例的多样本是指主管与员工这两群不同层级的受试者, 总计有 121 位主管与 215 位部属完成了一个简单的调查表。调查表中包括了测量组织承诺的九个题目, 经题项组合 (item parceling) 后产生三个组合测量变量 (co1、co2、co3), 以及员工绩效的三个指标变量: 工作满意度 (out1)、离职意愿 (out2) 与自我绩效评定 (out3)。这 6 个观察变量的描述统计量据列于表 10.5:

表 10.5 六个观察变量的描述统计量

变量	员工(N=215)		主管(N=121)		相关系数矩阵					
变量	M	SD	M	SD	OUT1	OUT2	OUT3	CO1	CO2	CO3
OUT1	4.44	.69	4.67	.57	1.00	.11	.39	.29	.22	.27
OUT2	2.62	1.40	2.07	1.22	-.35	1.00	.02	-.19	-.26	-.17
OUT3	4.40	.69	4.59	.59	.33	-.01	1.00	.15	.14	.25
CO1	9.29	1.83	10.04	1.64	.33	-.35	.03	1.00	.75	.67
CO2	9.26	1.92	10.05	1.50	.37	-.34	.08	.71	1.00	.78
CO3	8.99	1.76	9.75	1.55	.33	-.37	.05	.72	.79	1.00

注：相关系数矩阵中，下三角形部分为员工数据，上三角形部分为主管数据。

一、模型界定

从变量的关系来看，本范例的主要假设为“组织承诺感可以预测员工的工作表现”。其概念图示如图 10.2。从平均数的比较来看，本范例的另一个假设是“主管的组织承诺与工作表现水平与部属有所不同”。

由于平均数结构的检验必须建立在测量恒等性假设适当维系的前提之上。因此，下列两项假设必须逐一加以检验：

- 1. 测量模型是否具有跨样本（主管与员工）的恒等性？在此范例中，测量模式有两个，分别为外源测量变量的测量模型（组织承诺）与内生变量的测量模型（员工绩效表现）。
- 2. 因素平均数是否具有跨样本恒等性？此一假设还可区分为外源潜在变量（组织承诺）的平均数与内生潜在变量（员工绩效表现）的平均数是否具有跨样本的恒等性。

关于第一点，读者可以依据前一节的范例进行检测，在此不予示范。

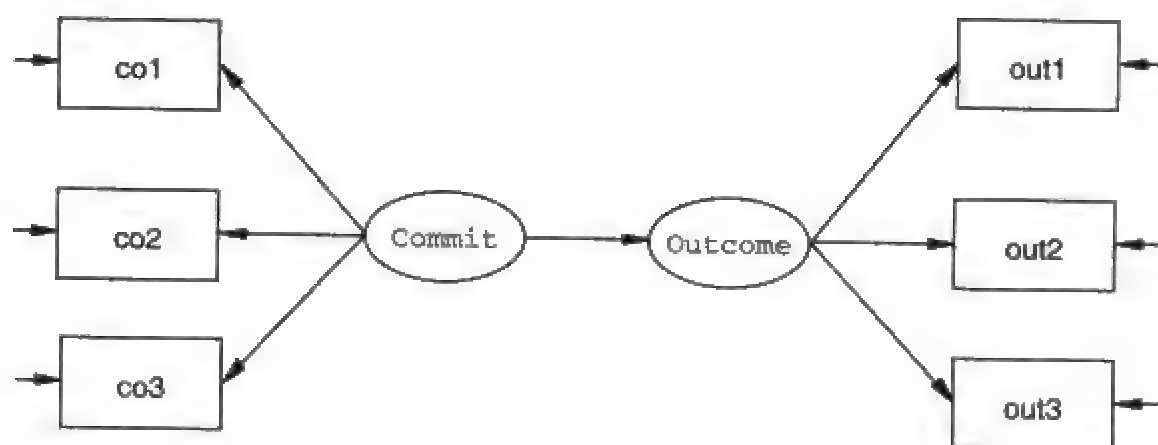


图 10.2 两个潜在变量的简单回归模型

从图 10.2 的路径图,可以推算出整个模型的测量数据数为 $(6 \times 7)/2 = 21$ ($DP=21$)。为了进行跨样本测量模型恒等性检验,必须增加测量模型的因素载荷、结构模型的回归系数被设定等同的条件(观察变量的测量残差可在两样本间自由估计,也可设定为等值)。此外,对于平均数结构的估计,另需增加两项设定:

1. 观察变量的截距被估计,共有 6 个 Tau 参数需被估计,但是这 6 个参数被设定为跨样本等同,因此两个样本仅需估计一次。
2. 潜在变量的平均数被估计,共计有 1 个 Kappa 参数与 1 个 Alpha 参数必须被估计。另一个样本的 Kappa 与 Alpha 参数被设定为 0,以作为参照组(reference group)。

二、LISREL 与 SIMPLIS 语法与参数估计

表 10.6 列出了 LISREL 与 SIMPLIS 语法。在 LISREL 语法部分,第 1 至第 11 行用以设定主管样本的模型,第 12 至第 16 行则用以设定员工样本的模型,两个样本的因素结构完全相同。输入的数据为原始数据(分别是 Ch10bm.dat 与 Ch10be.dat)。平均数结构估计需要两样本各观察变量上的平均数数据,建议使用原始数据库来读入数据。

平均数结构的设定位于第 5 行, TX=FR 与 TY=FR 表示主管样本的外源测量变

量与内生测量变量截距自由估计，AL=FR 与 KA=FR 表示主管样本的外源与内生潜在变量的平均数进行估计。

恒等性设定位于第 15 行的 LX=IN、TX=IN、LY=IN、TY=IN、GA=IN，表示因素载荷、测量变量截距、回归系数均具有跨样本的恒等性。AL=FI、KA=FI 表示员工样本的潜在变量平均数设定为 0，用来与主管样本进行平均数比较。加粗的部分若移除后，表示没有平均数估计的设定，在本范例中视为基准模型。

表 10.6 平均数结构分析的语法

LISREL 语法(Ch10.ls8)	
1	Group: Manager
2	DATA NI=6 NO=121 NG=2
3	RA FI=Ch10bm.dat
4	LA; out1 out2 out3 co1 co2 co3
5	MO NY=3 NE=1 NK=1 NX=3 PS=DI,FR PH=SY,FR GA=FU,FR TY=FR TX=FR AL=FR KA=FR
6	LE; Outcome
7	LK; Commit
8	FR LX(2,1) LX(3,1) LY(2,1) LY(3,1)
9	FI LX(1,1) LY(1,1)
10	VA 1 LX(1,1) LY(1,1)
11	OUTPUT SE TV SS SC MI
12	Group: Employee
13	DATA NI=6 NO=215
14	RA FI=Ch10be.dat
15	MO LX=IN LY=IN GA=IN TD=SP TE=SP TX=IN TY=IN AL=FI KA=FI
16	PD
17	OUTPUT SE TV SS SC MI
18	End of Problem
SIMPLIS 语法(Ch10b.spl)	
1	Group: Manager
2	Observed variables: out1 out2 out3 co1 co2 co3

```

3    raw data from file Ch10bm.dat
4    Latent Variables:  Outcome Commit
5    Relationships:
6    out1 = CONST 1.00*Outcome
7    out2-out3 = CONST Outcome
8    co1 = CONST 1.00*Commit
9    co2-co3 = CONST Commit
10   Outcome = Commit
11   Path Diagram
12   Group: Employee
13   Observed variables: out1 out2 out3 co1 co2 co3
14   raw data from file Ch10be.dat
15   Latent Variables:  Outcome Commit
16   Relationships:
17   Outcome=CONST
18   Commit=CONST
19   Set the Variance of Commit Free
20   Set the Error Variance of Outcome Free
21   Set the Error Variance of out1-co3 free
22   End of Problem

```

在 SIMPLIS 语法部分, 平均数结构的设定在测量模式位于第 6 至第 9 行, 各行均增加了 **CONST** 即表示截距纳入估计。以第 6 行为例, **out1=CONST 1.00*Outcome** 表示 out1 变量的截距需要估计, 但因为此题作为 Outcome 潜在变量的量尺化设定参数, 因素载荷设定为 1.00。

结构模式的潜在平均数设定位于第 17 至第 18 行。**Outcome=CONST** 表示员工样本的 OUTCOME 变量的截距需要估计, 而主管样本的截距则设定为 0, 作为参照。第 19 至第 21 行则表示残差方差与外源变量方差均设定为自由估计。模型中设定跨样本等同的则为因素载荷及回归系数, 因为在下半段的语法中没有出现这些设定, 表示套用主管样本的估计值。

三、估计结果与模型拟合

从报表中, 我们可以看到两个样本的共变量矩阵与观察变量的平均数。上半部是主管样本的共变量矩阵与各观察变量的平均数数据, 下半部则是员工样本的共变量矩阵与各观察变量的平均数数据。读者可以比对这些平均数数值是否与表 10.5 的描述统计量据一致。

Group: Manager

Covariance Matrix

	out1	out2	out3	col	co2	co3
out1	0.32					
out2	-0.08	1.48				
out3	0.13	0.01	0.34			
col	0.27	-0.38	0.14	2.67		
co2	0.19	-0.48	0.12	1.84	2.26	
co3	0.24	-0.33	0.23	1.69	1.81	2.40

Means

	out1	out2	out3	col	co2	co3
	4.67	2.07	4.59	10.04	10.05	9.75

Group: Employee

Covariance Matrix

	out1	out2	out3	col	co2	co3
out1	0.47					
out2	-0.33	1.95				
out3	0.15	-0.01	0.48			
col	0.41	-0.90	0.04	3.35		
co2	0.49	-0.91	0.11	2.49	3.68	
co3	0.40	-0.91	0.06	2.33	2.66	3.09

Means

	out1	out2	out3	col	co2	co3
	4.44	2.62	4.40	9.29	9.26	8.99

(一) 模型界定

模型界定报表列出了 31 个被估计的参数。在主管样本中, 被估计的参数包括了两个因素的方差 (6、7)、三个内生变量的测量残差变量的方差 (8、9、10)、三个外源变量的测量残差 (11、12、13)。平均数的估计则出现在 Alpha (20) 与 Kappa (21) (注 1)。

Group: Manager

Parameter Specifications

LAMBDA-Y EQUALS LAMBDA-Y IN THE FOLLOWING GROUP

LAMBDA-X EQUALS LAMBDA-X IN THE FOLLOWING GROUP

GAMMA EQUALS GAMMA IN THE FOLLOWING GROUP

PHI

Commit

6

PSI

Outcome

7

THETA-EPS

out1

8

out2

9

out3

10

THETA-DELTA

col

11

co2

12

co3

13

TAU-Y EQUALS TAU-Y IN THE FOLLOWING GROUP

TAU-X EQUALS TAU-X IN THE FOLLOWING GROUP

ALPHA

Outcome

20

KAPPA

Commit

21

注 1: 潜在变量的截距。Alpha 为内生潜在变量截距, Kappa 为外源潜在变量截距, 均反映变量的得分水平高低 (平均数)。

在员工样本部分，被估计的参数包括内生与外源测量模型的因素载荷参数（1至4），两个潜在变量的回归系数参数则位于第5个参数，各测量变量的6个截距参数位于第14至第19个参数（注2），这11个参数在两个样本间仅估计一次，因为这些参数在两个样本间已被设定为相等。

TAU-Y		
VAR 1	VAR 2	VAR 3
-----	-----	-----
14	15	16
TAU-X		
VAR 4	VAR 5	VAR 6
-----	-----	-----
17	18	19

注 2：测量变量截距。Tau-Y 为内生测量变量截距，Tau-X 为外源测量变量截距，皆表示测量变量得分水平高低（平均数）。

(二) 参数估计结果

平均数结构的路径分析，与 CFA 分析相似，最重要的参数是潜在变量平均数（Alpha、Kappa 参数）的估计，利用这些参数估计数的标准误，则可以计算出显著性检验的统计量（t 值）（注 3）。

ALPHA	
Outcome	

0.16	
(0.06)	
2.59	
KAPPA	
Commit	

0.73	
(0.17)	
4.40	

注 3：潜在变量的平均数估计数，以员工样本（潜在变量平均数为 0）作为基准值，所计算出来的跨样本差异统计量。正值表示主管高于员工样本。t 值大于 1.96 表示达到.05 显著水平，大于 2.58 表示达到.01 显著水平。

值得注意的是，本范例的 Alpha 与 Kappa 系数，是以主管样本来进行估计，员工样本的 Alpha 与 Kappa 系数被设定为 0，以作为与主管样本平均数的参照。当数值为正值时，表示主管样本的平均数高于员工。估计的结果发现，Commit 的截距估计数为.73（t=4.40，p<.01），Outcome 的截距估计数为.16（t=2.59，p<.01），两者估计数值均为正值，表示主管得分高于员工，且皆达.01 的显著水平。

(三) 模型拟合度分析

因为平均数结构检验是基于因素恒等性存在的情况下来进行测量变量截距估计与潜在变量平均数估计,所以,基准模型应为无平均数结构设定的恒等模型。由此,我们可以直接将表 10.7 之 LISREL 语法中,有关截距部分的设定均予以移除后,进行无平均数结构的恒等模型估计(语法文件为 Ch10bbase),得到基准模型的模型拟合度分析,结果列举于表 10.7:

表 10.7 平均数结构 PA-LV 路径分析模型拟合度评估摘要表

Model	WLS χ^2	df	RMSEA	NNFI	CFI	GFI	SRMR
基准模型	48.18 ($p=.00$)	21	.006	.96	.97	.96	.088
平均数模型	52.07 ($p=.00$)	25	.081	.96	.97	.96	.077
	$\Delta\chi^2=3.89$ ($p>.05$)	$\Delta=4$					

表 10.7 的数据显示,两个统合模型的拟合指数都十分接近,而两者的卡方值的差异量未达显著水平($\Delta\chi^2=3.89$, $\Delta df=4$, $p>.05$),表示增加平均数与截距参数并不影响模型拟合度的计算。但是值得注意的是,由于增加截距与平均数参数并未改变模型结构,仅增加自由参数的数目,因此卡方差异量单纯反映了 SEM 技术复杂度提高后的影响,而和假设模型与实际观察数据的拟合与否无关。

(四) 完成 SEM 分析

终解的数据采用跨样本的完全标准化解,也就是 COMMON METRIC COMPLETELY STANDARDIZED SOLUTION,以符合两样本具有跨样本等同性的基本假设的精神。因素载荷(注 4)、回归系数(注 5)均是重要的参数,整理后的估计结果与显著性检验结果列于表 10.8。

Group: Manager

Common Metric Completely Standardized Solution

LAMBDA-Y	
	Outcome

out1	0.67
out2	-0.47
out3	0.34
LAMBDA-X	
	Commit

co1	0.81
co2	0.89
co3	0.88

注 4：参数估计的跨样本完全标准化解。有负值表示该题与潜在分数的高低呈现反方向关系。

GAMMA	
	Commit

Outcome	0.56

注 5：组织承诺对于员工绩效的回归系数。

Covariance Matrix of ETA and KSI

	Outcome	Commit

Outcome	0.64	
Commit	0.44	0.78

PSI	
	Outcome

	0.39

THETA-EPS		
out1	out2	out3
-----	-----	-----
0.54	0.74	0.69

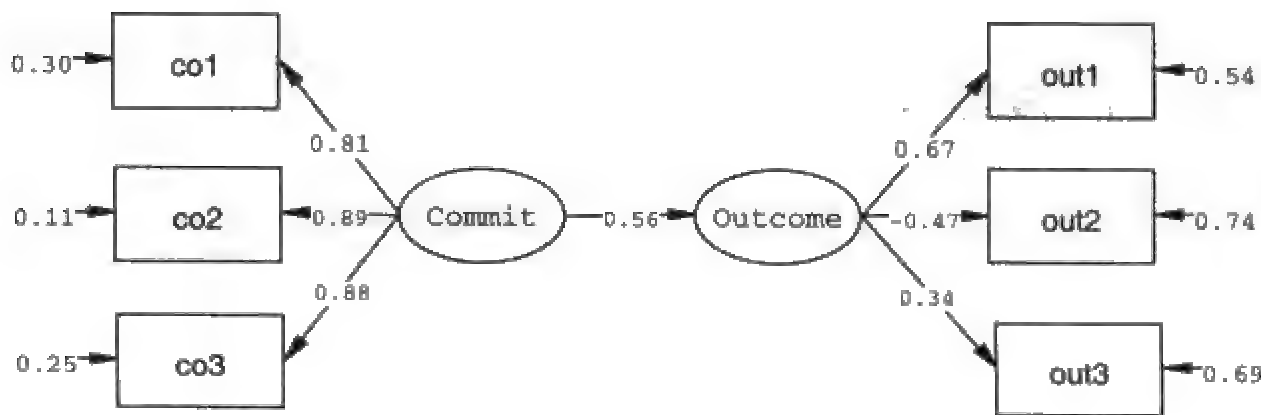
THETA-DELTA		
col	co2	co3
-----	-----	-----
0.30	0.11	0.25

表 10.8 的数据清楚地说明了两个潜在因素平均数的组间差异，主管样本的组织承诺与绩效均较员工样本为高，*t* 检验均达.01 统计水平，标准化回归系数为.56，表示组织承诺可以有效预测绩效，整体的拟合度大致良好 ($\Delta\chi^2=52.07$, $df=25$, $p=.00$, RMSEA=.081)，标准化终解的数据（图 10.3）以及平均数结构估计数（标准化前与标准化后）如图 10.4 所示。

表 10.8 统合模型的平均数结构参数估计结果整理表

变量		因素载 荷	残差	截距		因素	因素平均数		t
				原始	标准化		原始	标准化	
测量模型						结构模型			
外源 变量	CO1	.81	.30	9.32	5.26	组织承诺	.73	1.04	4.40**
	CO2	.89	.11	9.25	5.21				
	CO3	.88	.25	8.95	5.33				
内生 变量	OUT1	.67	.54	4.42	6.90	员工绩效	.16	.07	2.59**
	OUT2	-.47	.74	2.60	1.91				
	OUT3	.34	.69	4.41	6.71				
回归系数							.17	.56	7.34***

** $p<.01$ *** $p<.001$



Chi-Square=52.07, df=25, P-value=0.00118, RMSEA=0.081

图 10.3 统合模型最终解（跨量尺完全标准化解）图示

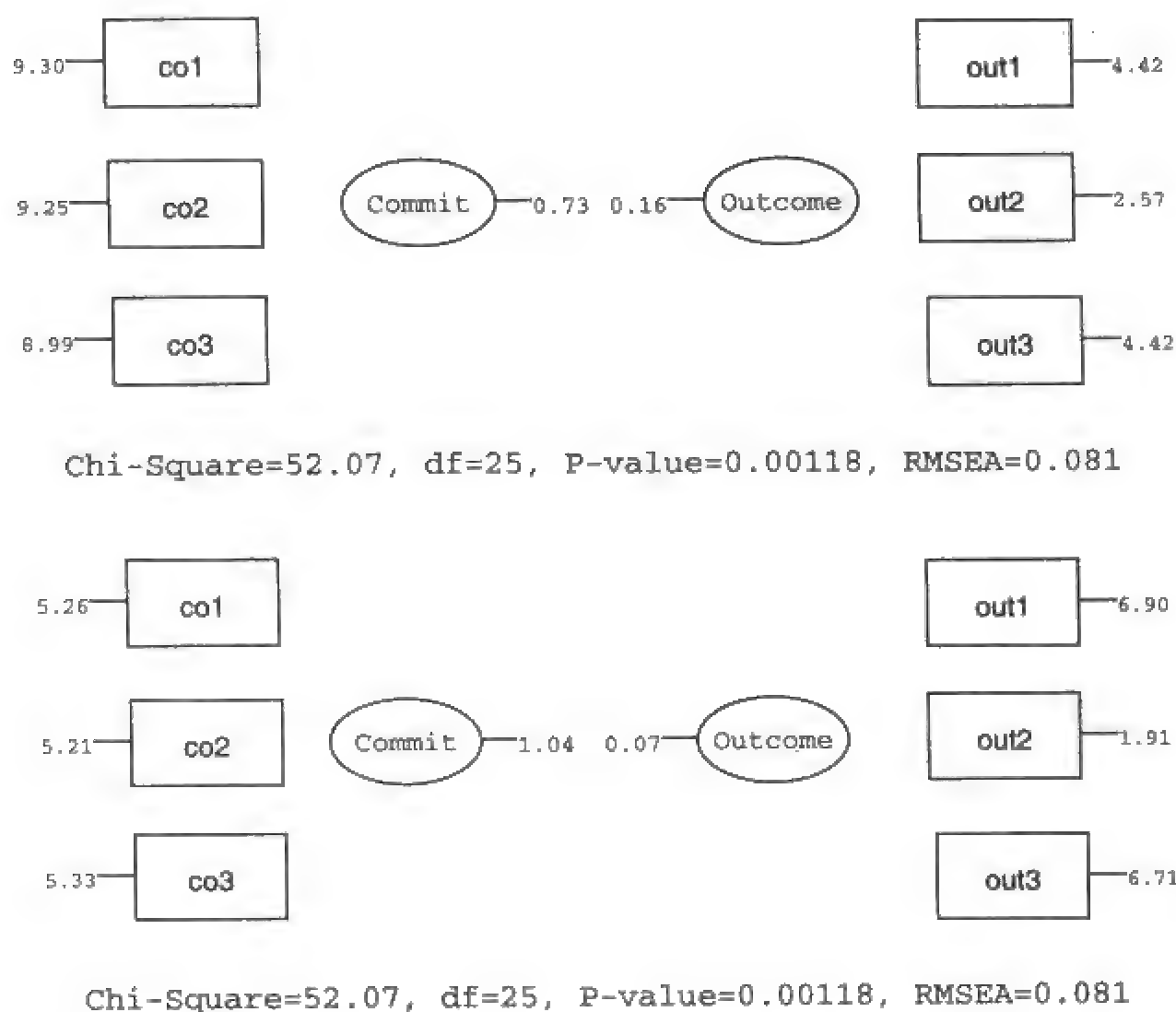


图 10.4 标准化前（上）与标准化后（下）平均数估计结果图示

四、AMOS 平均数结构的示范

统合模型的 AMOS 平均数结构分析，包括了内生与外源潜在变量平均数的估计，说明如下。

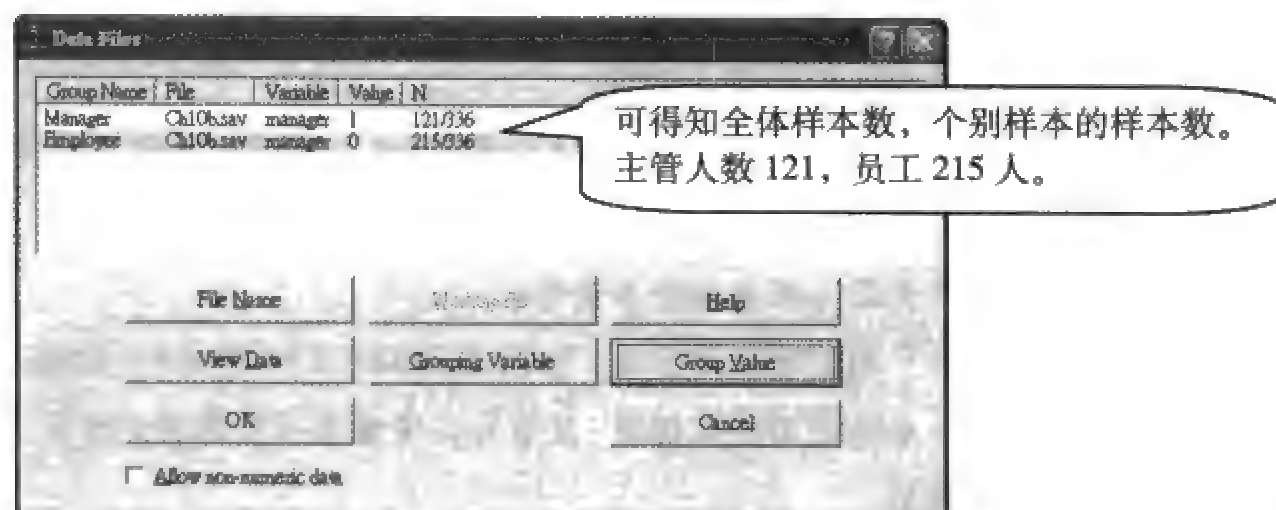
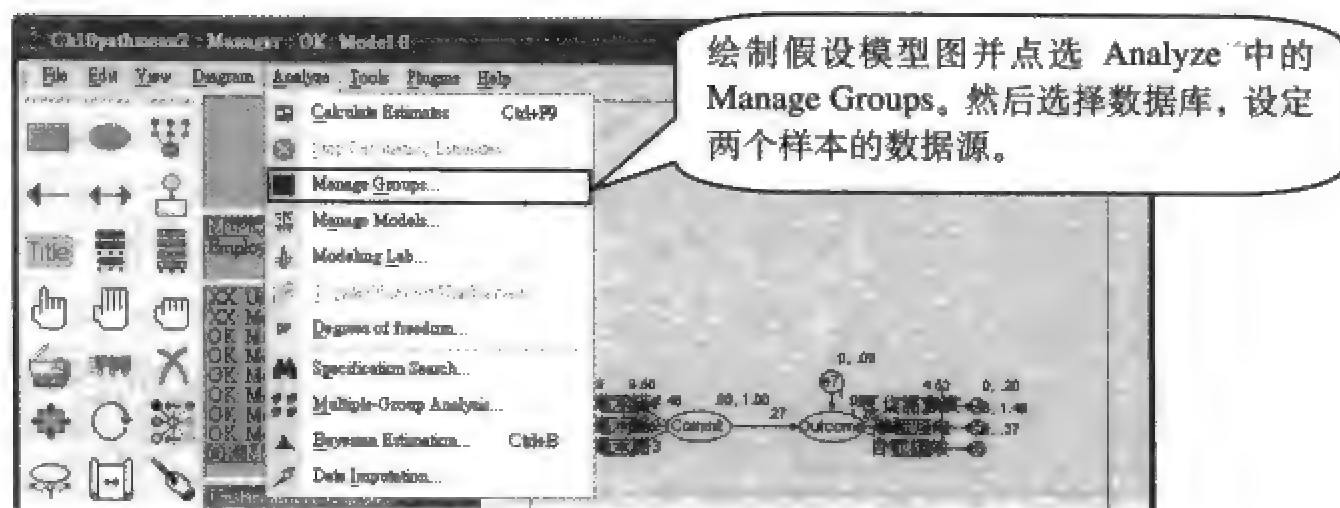
（一）操作步骤

步骤一：绘制假设模型图，选择数据文件，完成模型的设定


步骤二：點選 Analyze->Manage Groups

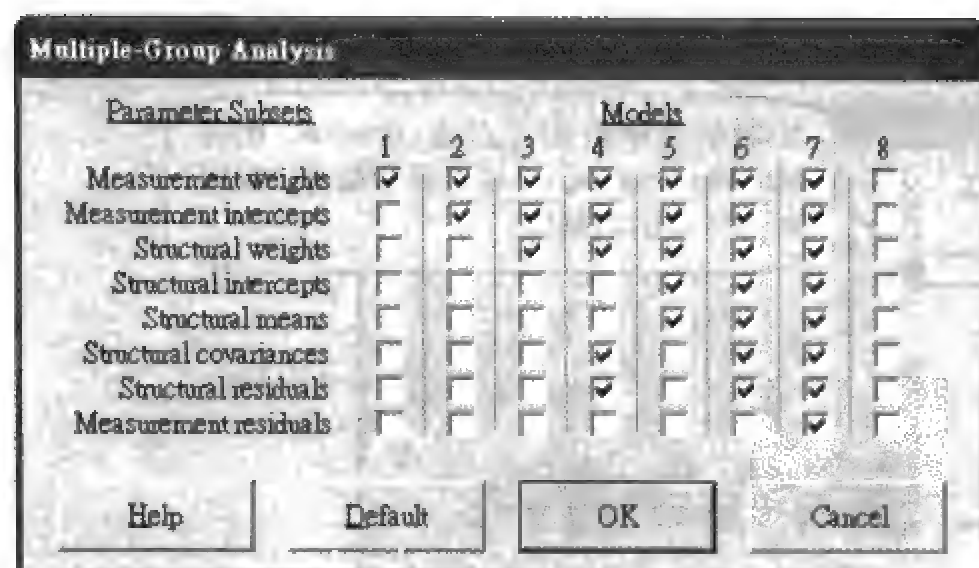
步骤三：设定样本名称：Manage 与 Employee

步骤四：点选数据文件，开启档案，设定两个样本的来源



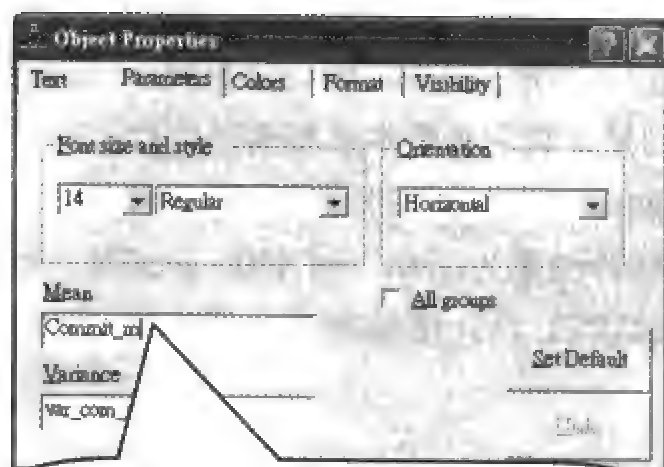
步骤五：点选分析选项，点选 Estimate means and intercepts，宣告要进行截距与平均数的估计。

步骤六：点选  (多样本分析)，按确定之后即进入模型界定对话框如下（前四个模型的结构平均数与截距自由估计，后三个则无估计）：

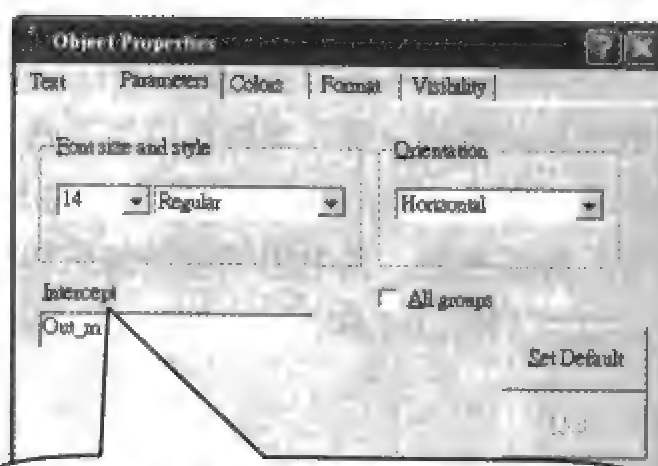


1. 第一个模型为因素载荷跨样本等值。
2. 第二个模型增加了观察变量截距跨样本恒等。
3. 第三个模型增加了结构参数（回归系数）跨样本恒等。（与 LISREL 范例相同设定之模型）。
4. 第四个模型增加了结构共变量与残差跨样本恒等。
5. 第五个模型令结构截距与平均数跨样本恒等，用来与 Model 3 和 Model 4 比较。
6. 第六个模型令结构截距与平均数跨样本恒等，且结构共变量与残差跨样本恒等。
7. 第七个模型为完全等值模型。

步骤七：设定其中一个样本（本范例为主管样本）的外源潜在变量平均数为自由估计，内生潜在变量截距为自由估计。（首先点选样本为 Manager 后，将鼠标移至将鼠标光标移至两个因素上，按压右键后，选取 Object Properties 即可获得下列画面。）



主管样本的外源潜在变量平均数自由估计，给定一个参数为 Commit_m。因素方差自由估计。



主管样本的内生潜在变量截距为自由估计，给定一个参数为 Out_m。

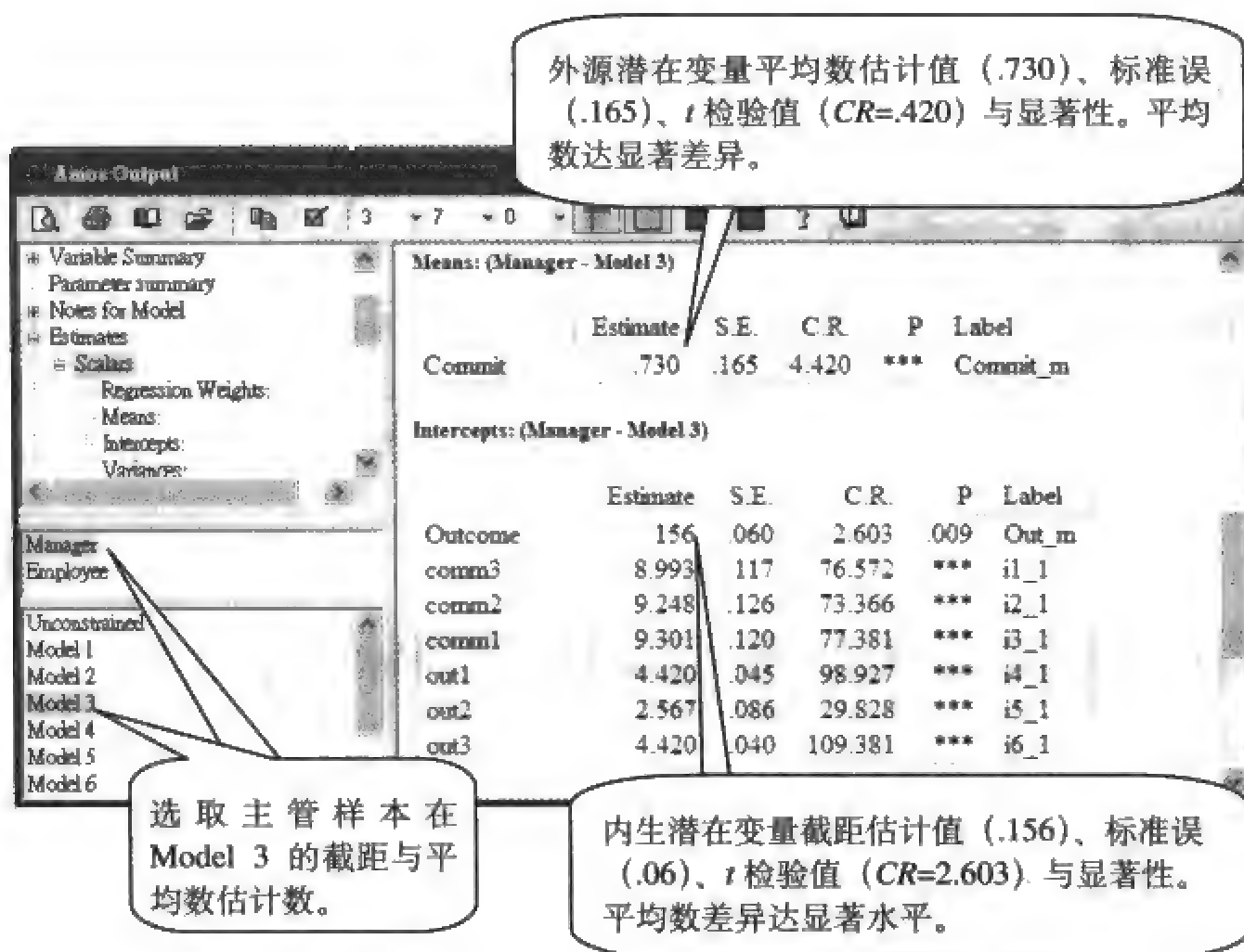
(二) 估计结果：模型拟合比较

进行估计后，AMOS 将会将六个模型（包含未设限模型、饱和模型与独立模型）的估计结果列于报表中。以下是模型拟合指数。报表中，Model 3 是比照前面 LISREL 范例的设定（观察变量常态与截距为跨样本等值）， $\chi^2=49.491$ ， $df=25$ ， $p<.001$ ，结果与先前 LISREL 估计结果相似。

Model Fit Summary						
CMIN						
Model	NPAR	CMIN	DF	P	CMIN/DF	
Model 2	30	48.534	24	.002	2.022	
Model 3	29	49.491	25	.002	1.980	
Model 4	27	56.116	27	.001	2.078	
Model 5	27	76.413	27	.000	2.830	
Model 6	25	80.855	29	.000	2.788	
Model 7	19	99.762	35	.000	2.850	
Saturated model	54	.000	0			
Independence model	24	763.748	30	.000	25.458	

(三) 估计结果：平均数差异检验

平均数差异的检验主要并非检视模型拟合，而是比较两个样本在结构平均数上的差异是否具有统计意义。可以检视估计数当中的 Means 即可在结构平均数设定为自由估计样本（主管）得到报表。其中组织承诺的平均数估计结果为.730， $t=4.420$ ， $p<.001$ ，绩效截距估计数为.156， $t=2.603$ ， $p<.001$ ，与先前 LISREL 估计结果相近。



五、Mplus 估计语法与结果

(一) Mplus 语法

TITLE: Ch10b Mean structure modeling
DATA:

NGROUPS=2;

FILE (manager) IS Ch10bm.dat;

FILE (employee) IS Ch10be.dat;

VARIABLE: NAMES ARE out1-out3 col-co3;

MODEL:

Commit by col-co3;

Outcome by out1-out3;

Outcome on Commit(1);

OUTPUT:

STANDARDIZED;

多样本分析的设置格式

NGGROUP=2 表示有两个样本, 数据各自存放于两个档案中。

模式设定

Commit 由 col-co3 三个变量估计

Outcome 由 out1-out3 三个变量估计

两个样本套用相同的因素负荷量、观察变量截距

回归系数恒等设定

Commit(1)括号中数值表示设定为跨样本等值

(二) Mplus 报表

SUMMARY OF ANALYSIS

Number of groups	2
Number of observations	
Group MANAGER	121
Group EMPLOYEE	215

表示两个数据库的状况
各有 121 与 215 笔资料

Number of dependent variables	6
Number of independent variables	0
Number of continuous latent variables	2

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

Value	49.787
Degrees of Freedom	25
P-Value	0.0023

模型拟合数据

卡方值为 49.787, 与 LISREL 估计值相近

Chi-Square Contributions From Each Group

MANAGER	24.095
EMPLOYEE	25.692

两个样本各自估计的模型拟合指针卡方值, 分别合计即为总和的 49.787。

Chi-Square Test of Model Fit for the Baseline Model

Value	768.322
Degrees of Freedom	30
P-Value	0.0000

CFI/TLI

CFI	0.966
TLI	0.960

Loglikelihood

H0 Value	-2850.338
H1 Value	-2825.444

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.077
90 Percent C.I.	0.045 0.108

SRMR (Standardized Root Mean Square Residual)

Value	0.075
-------	-------

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Group MANAGER				
COMMIT BY				
CO1	1.000	0.000	999.000	999.000
CO2	1.106	0.058	19.101	0.000
CO3	1.034	0.055	18.827	0.000
OUTCOME BY				
OUT1	1.000	0.000	999.000	999.000
OUT2	-1.466	0.390	-3.756	0.000
OUT3	0.528	0.103	5.137	0.000
OUTCOME ON COMMIT	0.169	0.023	7.207	0.000
Means				
COMMIT	0.000	0.000	999.000	999.000
Intercepts				
OUT1	4.699			0.000
OUT2	2.157			0.000
OUT3	4.568	0.042	109.443	0.000
CO1	10.030	0.132	75.965	0.000
CO2	10.035	0.135	74.553	0.000
CO3	9.748			0.000
OUTCOME	0.000			9.000

因素载荷量与回归系数均设定为两样本等值。

观察变量的截距
亦为跨样本等值

潜在变量的截距在第一
个样本设定为 0

Variances				
COMMIT	1.583	0.257	6.160	0.000

Residual Variances				
OUT1	0.218	0.057	3.844	0.000
OUT2	1.319	0.194	6.789	0.000
OUT3	0.297	0.045	6.674	0.000
CO1	0.943	0.147	6.405	0.000
CO2	0.331	0.101	3.272	0.001
CO3	0.697	0.123	5.649	0.000
OUTCOME	0.071	0.048	1.473	0.141

Group EMPLOYEE

COMMIT BY				
CO1	1.000	0.000	999.000	999.000
CO2	1.106	0.058	19.101	0.000
CO3	1.034	0.055	18.827	0.000

OUTCOME BY				
OUT1	1.000	0.000	999.000	999.000
OUT2	-1.466	0.390	-3.756	0.000
OUT3	0.528	0.103	5.137	0.000

OUTCOME ON COMMIT				
	0.169	0.023	7.207	0.000

Means				
COMMIT	-0.730	0.165	-4.430	0.000

Intercepts				
OUT1	4.699			
OUT2	2.157			
OUT3	4.568			
CO1	10.030			
CO2	10.055			
CO3	9.748			
OUTCOME	-0.156	0.132	74.557	0.000
		0.059	73.955	0.000
			-2.627	0.009

潜在变量的截距在第二个样本设定为自由估计。Commit 估计值为-.730, $t=-4.43$, Outcome 估计值为-.156, $t=-2.627$, 均达统计显著水平, 与 LISREL 结果相近。

Variances				
COMMIT	2.293	0.296	7.740	0.000

Residual Variances				
OUT1	0.228	0.065	3.500	0.000
OUT2	1.449	0.199	7.284	0.000
OUT3	0.426	0.046	9.268	0.000
CO1	1.136	0.139	8.146	0.000
CO2	0.816	0.129	6.324	0.000
CO3	0.627	0.107	5.853	0.000
OUTCOME	0.155	0.065	2.402	0.016

STDYX Standardization

标准化估计数报表

值得注意的是, Mplus 并不提供跨样本完全标准化解, 因此两个样本的标准化解并不一样。这里的标准化解等同于 LISREL 的 Within Group Completely Standardized Solution

Est

Group MANAGER

COMMIT	BY				
CO1		0.792	0.035	22.639	0.000
CO2		0.924	0.024	37.942	0.000
CO3		0.842	0.030	28.058	0.000
OUTCOME	BY				
OUT1		0.589	0.127	4.652	0.000
OUT2		-0.399	0.076	-5.282	0.000
OUT3		0.313	0.095	3.310	0.001
OUTCOME	ON				
COMMIT		0.623	0.135	4.612	0.000
Means					
COMMIT		0.000	0.000	999.000	999.000
Intercepts					
OUT1		8.121	0.488	16.640	0.000
OUT2		1.722	0.138	12.507	0.000
OUT3		7.954	0.495	16.079	0.000
CO1		6.311	0.359	17.561	0.000
CO2		6.679	0.430	15.548	0.000
CO3		6.305	0.378	16.683	0.000
OUTCOME		0.000	0.000	999.000	999.000
Variances					
COMMIT		1.000	0.000	999.000	999.000
Residual Variances					
OUT1		0.652	0.149	4.368	0.000
..... (略)					

第五节 结 语

本章内容介绍了多样本结构方程模型分析应用于平均数结构的检验。由于 SEM 模型相当复杂,除了因素结构本身的复杂性,还包括结构模型的设定,因此 SEM 分析估计出来的截距或平均数数据虽然可以反映差异的有无,但是造成差异的原因却无法从 SEM 分析的报表中获得。由此可见,样本本身性质的描述,即成为多样本分析成败的主要关键所在。例如各样本的差异是否可以具体描述,除了某些明显的差异之外,其他性质是否维持一致,如果可以,从 SEM 分析获得的样本间差异,即可归因于样本间差异的明显特质。相对的,当不同样本之间的异质性如果太高,我们即很难理解 SEM 分析得到的样本间差异的原因究竟为何。

样本异质性的问题特别容易发生在非随机抽样所获得的样本,例如某个班级的男生与女生样本,某个公司的主管与员工。在社会与行为科学领域,除非进行实验室研究,可以将样本利用随机分派区分成性质完全相等的两组,否则两个独立样本之间很难具有高度的方差齐性。如此一来即造成样本间比较的困难度。因此,在进行多样本模型分析时,研究者宜特别注意抽样的问题,以免因为样本本身的因素造成分析或解释上的偏差。

第十一章

结构方程模型的正确运用

第一节 正确运用 SEM 的相关议题

统计技术的发展在 1960 年可以视为一个分水岭。之前的心理计量学的发展以古典的推论统计技术为主,例如 t -test、ANOVA、Pearson's r 、least square regression;之后的发展则着重于各种统计问题的校正技术,可称之为现代统计学(modern statistics),例如 M statistics (trimmed mean statistics)、percentage-bend correlation、Theil-Sen estimator 等。1994 年当《Structural Equation Modeling》期刊第一卷出版后,可以称为统计技术的第三波革命。但是尽管 SEM 的发展风起云涌,但是同时也遭遇到许多的困境与挑战。例如美国南加州大学(University of Southern California)的心理系教授 Norman Cliff 在 1983 年以相当严峻的口吻质疑 SEM 的不当使用:第一,研究者所获得的数据无法替我们完全确认或否认一个模型的正确性,因为模型是人为的,而且可以各种方法重新定义;第二,具有时间性的先后次序证据并不代表因果;第三,潜在变量的命名是一个主观的历程,而非客观的事实,潜在变量的估计存在着名义谬误(nominalistic fallacy)的陷阱;第四,事后的解释与调整具有诚信与可信度的问题,也就是驳斥部分研究者大量使用模型修饰程序获得理想拟合度的不当做法。

Cliff 的批评可以说是一种警世之语。事实上,Cliff 也相当看重 SEM 的发展,他在 1983 年的文章中最后提到:

“.....最后,我必须再次强调,像 LISREL 这类的分析工具的确提供了一个空前的、史无前例的机会,使我们能够把这类的研究好好地做好。”

虽然 SEM 是一套非常具有说服力的分析技术与研究范式,但是 SEM 本身仍存在一些非常基本的问题,这些不论是 SEM 本身的技术问题,或是 SEM 应用上的限制,一再影响着 SEM 分析结果的正确运用。

一、SEM 运用的三个关键议题

Nesselroade (1991) 指出,使用 SEM 这一类高阶的复变量分析技术,有三个层面的问题必须注意:个体(individuals)、测量(measures)、情境(occasions)。首先,

与个体有关的问题，也就是与抽样与样本有关的问题。任何推论统计，都涉及了抽样的过程。因此，抽样与样本对于统计分析的影响，在任何一本统计著作皆有详细的说明。一般对于统计稍有概念的研究者，对于抽样的原理与相关定理的重要性也应知之甚详。

对于 SEM 而言，由于分析的核心原理是比较理论值与实际值的差异程度，因此，SEM 深受样本与样本统计量的影响。尽管 SEM 发展了诸多的估计样本所带来的影响分析策略（例如复核效化）与统计指数（例如 ECVI），但是这些技术或指数也有自己的限制，更重要的是无法改变抽样本身会带来各种问题的事实。因此，对于抽样程序与样本特性可以说是 SEM 成败的关键之一。

其次，测量的问题也深深影响着 SEM 的操作。所谓测量的问题主要在于潜在变量的估计与分析。不论是古典测验理论对于心理测验的界定，或是因素分析统计技术对于因素的决定，都反映了心理计量学最重要的任务之一，是能够有效处理抽象心理构念的测量问题。对于潜在的心理构念（construct），我们可以视其为一个行为或心理特征的概念总体（population），我们所进行的测量，是从这个概念总体抽取相当的样本指标来反映这个总体概念。因此，测量的成败取决于我们如何对于这个测量总体进行定义。一旦对于总体加以定义之后，我们才可能发展出相对应的一组样本指标来反映该总体的基本特征。

如果从前面一段所谈到的抽样问题来看，从概念总体选取指标样本，就好比从一群人口总体中抽取样本进行研究，有异曲同工之意。但是，不同的是人口总体通常并没有太困难的定义问题，因为人的集合是具体的事件。然而对于心理特质的定义，则是一个相当抽象的程序。这些有关于测量题目如何选取、藉以反映潜在变量性质以及不同的选题程序对于 SEM 影响的选择效应问题（selection effects；Maccallum & Austin, 2000），是 SEM 成败的第二个关键点。

最后，SEM 使用的情境也是 SEM 分析的重要影响因素。举例来说，当 SEM 应用于重复量数的研究数据分析时，不同测量时间点的重复测量的操作方式，使得前面提到的样本与测量问题之外，又增加了一个不同情境的干扰效果。在近年来获得相当发展的成长曲线分析的应用上，对于不同时间点下的 SEM 测量应如何进行有着诸多的讨论，反映了这个议题的重要性。此外，复核效化的应用也说明了 SEM 的结果在不同人口总体的类化性，也是一个必须加以探讨的重要议题。

进一步的，本书并没有特别讨论的多重特质多重方法矩阵（MTMM design）研

究的 SEM 分析,这类应用也涉及了 SEM 模型在多元测量模式当中的应用问题。如果我们把多重方法视为 SEM 分析的不同情境(好比重复测量的不同时间点),MTMM 设计的应用也反映了情境影响 SEM 分析这个议题的重要性(参见 Eid, Lischetzke, Nussbeck, & Trierweiler, 2003),因为时间序列分析、复核效化以及 MTMM 设计的应用,都是 SEM 最近在期刊上讨论非常密切的课题。我们或许不应该说情境议题是 SEM 分析成败的另一个关键所在,而是将这一点视为 SEM 更高度发展的跳板因素。对于这些议题的发展,读者可以拭目以待。

二、SEM 的决策建议

从原理来看,SEM 是一套非常复杂的统计技术,但是由于应用程序的便捷与强大功能,使得用户可以不需对于 SEM 的原理与内涵有深入了解,即可以快速熟练操作 SEM,得到大量的数据,同时窗口版的 LISREL、EQS、AMOS 等 SEM 分析工具更可以结合绘图功能与文书作业软件,使用者甚至不必绘制变量关系的路径图,直接从软件中制作。从 SEM 的推广来看,这些便捷的作业模式并不令人鼓舞。

综观统计技术的运用,从初等统计、多变量统计到结构方程模型,每一种统计方法都难免必须进行主观的判断。尤其是当统计的原理越趋于繁复时,通常所涉及的人为判断更多且决策更为困难。基本上,操作 SEM 具有一些值得注意之处,虽然在本书的各章节与范例当中已有讨论,以下则列举了十点有关 SEM 分析决策上的提醒与建议,系综合各个层面的概念汇整而成,供读者们参考。

1. 对于模型的优劣应该避免武断与绝对的结论。记得在单一的 SEM 分析中,我们只能证实某一个模型可以拟合观察数据,也就是该模型没有错误地去代表观察值,但是我们并无法证明某一个模型是绝对正确的。
2. 对于所检验的模型皆以理论为基础。在没有确切的根据或没有逻辑推理的合理性之前,不要随意更动模型的界定。即使模型修饰指数告诉我们某些参数可以变动调整,但必须全盘考虑整个模型的设计原理,避免“为改善拟合度而修饰”的做法。
3. 尽可能检验不同的模型,了解不同模型的计量特性。如果可以的话,研究者应该从不同的理论观点去提出竞争的对立假设来进行相互比较;必要时,可

以进行复核效化的检验,利用不同的样本来重复检验,以确认模型估计的类化能力。

4. 当一个 SEM 模型当中兼含测量模型与结构模型时,宜先进行测量模型的检验,待测量模型具有相当的合理性之后,再进行结构模型参数估计。使得 SEM 模型评估具有测量的“渐进合理性”。
5. 对于模型界定的适当性,不应只考虑统计量据的意义,而需考虑概念的本质与意义。尤其应该注意潜在变量的多重指标是否具有理论与实务的适当性。
6. 对于模型拟合度的评估,应同时兼采各种指数,因为不同的指数反映不同模型的计量特征。同时,不论指数数据的好坏,均应翔实报告,不应“报喜不报忧”或“粉饰太平”。
7. 尽可能先行检验数据的分布特性,使其不受偏离值与遗漏值的影响。必要时应进行多变量的正态化或遗漏值检验,找出潜藏于变量之间的特殊数据。
8. 对于测量变量的选择,应尽可能符合正态化假设的基本特性,例如避免二分变量或顺序变量的虚拟化;单一测量变量的尺度尽可能的增加,并扩大其变异量(例如可以将多题集成单一测量变量,提高单一指标的变异量,使其符合正态化的特质)。
9. 当不同模型具有类似的拟合度时,使用简约原则决定最终接受的模型。也就是越简单的模型(估计的参数较少者),越具有较高的简约性。
10. 样本数越大越好。

第二节 SEM 的解释与应用

自从 SEM 分析的概念被提出以来,最大的争议与批评并不是在技术层面,而是在整个研究范式的应用与解释(Freedman, 1987; Cliff, 1983)上的争议。因此,在讨论如何解释 SEM 分析的结果之前,我们先必须澄清 SEM 分析在因果论证与结果推论上的一些限制。

一、因果关系的论证

在 SEM 的使用上,最常被提出讨论的是因果关系的质疑。SEM 最早的名称是因果模型分析 (causal modeling)。到目前为止,仍有很多 SEM 研究者将回归分析或是结构方程模型视为非实验研究可以得到因果结论的重要技术。然而,如果从研究设计的角度来看,只有经过严格实验处理的研究,将受试者经过随机分派,然后在严谨的实验控制下,检验研究者所操纵的自变量变化对于因变量的影响,此时所得到的结果才符合因果关系的基本条件;而其他任何形式的研究设计或统计推论,所得到的只可能是变量的预测或解释而非因果关系。

实验研究当中,研究者所操纵的自变量(因)是影响因变量(果)的直接原因,在时间顺序上有前因后果之分。但是,非实验研究的变量关系往往没有一个清楚的方向与单一轴线,变量的影响关系无法清楚区分。同时,在变量的选取上,通常是从因变量来回溯、探讨可能具有解释力的前置变量,再放入模型中检验研究者的推理是否正确。因此,如果以 SEM 分析的测量原理,要从同一个时间点下所测量得到的变量关系去推导出因果关系,是不切实际的。SEM 的统计原理与其他相关于回归分析的并无二致,在推导因果结论的限制是相同的 (Holland, 1986)。除非 SEM 分析所进行的研究在测量上能够证明变量的前因后果,(例如纵贯性研究,变量的测量是在不同时间点下的测量),此时关系的讨论或许可以被接受是一种因果的概念;否则,任何其他形式的 SEM 分析,都应该避免做出因果的结论¹。

二、SEM 分析的推论限制

结构方程模型是一套用以检验特定假设模型的统计方法学。因此,SEM 最主要的一个目的是在验证研究者所提出的理论或概念架构是否具有实证的意义。整个 SEM 的分析程序都离不开研究者所提出的 SEM 模型,因而研究者是否可以在提出研究问题的第一时间,就能够透过理论推导与文献检阅过程,选择适当的研究变量,

¹有关因果关系的论证, Pearl (2000)、Cliff (1983) 与 Mulaik (1986, 1995) 曾有详细的说明与精彩的论辩,有兴趣的读者可以参考他们的相关论文。

提出有意义的研究假设去说明变量的关系,进而发展出适切理想的假设模型,即成为 SEM 模式的计量检证程序是否可以顺利完成的一个基本条件。换句话说,SEM 分析技术只是一套统计的方法与分析的策略,SEM 的本身并无法创造理论或知识,而是需要研究者以其智慧去整理前人所发展出来的理论或知识,建构出一套适当的概念模型,然后再以 SEM 技术协助研究者完成模型的分析与讨论。

SEM 分析最大的禁忌是过度倚赖技术指标的数据与过度推论。SEM 分析只能用来评估研究者所提出的假设模型是否适切,但是,究竟何者才是真正能够反映真实世界的变量之间关系的模型,这一个结论并不能够从模型拟合指数的数据来得到。因为除了研究者所提出的模型之外,同样的一组观察变量可能有许多不同的模型组合,这些基于同样观察数据基础的假设模型可能都有理想的拟合度,SEM 分析并无法区辨这些计量特征类似的理论模式何者为真。SEM 的使用者不但必须谨记统计方法学本身的限制,也必须避免陷入过度推论的陷阱当中。

从等同模型比较的实际范例,我们即可以看出上面讨论的问题如何影响 SEM 研究的进行。Lee 和 Hershberger (1990) 与 MacCallum、Wegener、Uchino 和 Fabrigar (1993) 等学者就曾经证明两个不同的模型具有一样的自由度,但是分别用以描述两套不同的假设概念。此时,在统计上两个模型是等同的,但是确有两种不同的估计结果。或者简单来说,我们如果将一个 SEM 模型中,两个潜在变量的影响路径加以倒置,也就是把 A 指向 B 的参数,改为 B 指向 A 的参数,此时模型的自由度是一样的,模型的复杂度是一样的,变量是一样的。但是模型代表的意义却有极大的不同,更棘手的是,倒置后模型参数估计结果,可能与倒置前的估计数相差不多。此时,从技术面是无法解释此种差异的。唯一的解救途径,是回头找出理论上的解释支持哪一种模型的设计。

三、SEM 分析的解释

从先前的讨论中,我们可以意识到结构方程模型的分析有其限制,但是并不意味着 SEM 分析的结果缺乏说服力。只要依循 SEM 分析的原理,逐步完成各项设定与检验工作,我们还是可以从分析结果中得到许多具有启发性的研究发现。

首先,使用者必须了解 SEM 分析的各项参数估计与模型拟合度数据,是变量因果关系的必要但非充分的证据。SEM 分析的证据可以说明某一个因果概念是可能存在的,但是不能据以排斥其他模型的存在。除非 SEM 分析直接检验了其他假设模型的计量特性,得到清楚的结论,证明该模型是不适当的。否则,SEM 分析的结果应仅就该研究所使用的样本与所检验的变量关系来讨论。

此外,由于 SEM 分析结果受到诸多因素的影响,例如样本数的大小、变量正态性问题、估计方法的选择、潜在变量的量尺化设定等。在研究报告撰写的时候,也应详细说明这些技术层面的特殊之处与处理原则,尽可能使分析过程透明化。此举除了有利于其他读者的了解之外,也是让其他研究者得以复制研究的结果,维系一个 SEM 分析公正、客观的重要手段。

第三节 结语: SEM 的展望

要为 SEM 预期一个未来,就像替别人算命一样,是非常不容易的事情,但是要说明目前计量领域对于 SEM 的重视与未来发展的高度信心,却是非常容易的一件事。例如在心理学领域的重要文献《心理学年度评论》(*Annual Review of Psychology*)上,近年来陆续刊登与 SEM 有关的综述文献。一些重要期刊当中的 SEM 文章数量也不计其数,市面上可以用琳琅满目来形容 SEM 分析软件,以及 SEM 专门的书籍的陆续、密集出版,都表示 SEM 是一个发展成熟但仍有相当可以发挥的测量、统计、方法学领域。

SEM 应用在亚洲各国于 2000 年之后逐渐蓬勃,但从专门著作、研究者社群、专属期刊、相关工作坊、大专院校的开课列表,结构方程模型的影子仍属稀有,显示 SEM 在亚洲地区仍是大有可为。我会抱持乐观的观点,理由有三:第一,如果以 SEM 受到国际间学术界的重视程度以及国内学术界追求国际化的趋势来看,透过 SEM 的语言,参与国际间的对话是一个非常值得鼓励的做法。第二,国内已有越来越多的论文投稿,开始使用 SEM,尤其是年轻一代的学者,对于西方学术发展接触程度高,对于 SEM 都十分愿意尝试。显见学术界对于 SEM 的需求与潜在使用者有增加的趋势。

第三,以国内整体学术发展趋势与信息水平来看,十分有利于 SEM 的发展,尤

其是统计分析技术的计算机化,已经相当普及。部分大学教育与心理学系教授统计学或测验学的老师,已经开始采用 SPSS、SAS、STATA 软件来协助教学;另外,几乎所有的研究生论文,都运用 MINITAB、SPSS 或 SAS 等统计软件进行数据分析,对于 LISREL 或 EQS、AMOS 的推广,可以说是水到渠成。但是,要使这个乐观的期待能够实现最重要的是需要推手,将 SEM 推上国内学术界的舞台。本书出版的主要动机,也就是在促成此一理想的实现。

本书分量不薄,希望能够从 SEM 的基本概念、原理、操作技术,到应用的范式,完整地进行解说与示范,使读者能藉由此书的阅读与范例的练习熟悉 SEM 这套统计方法学。本书的撰写,另一方面也是配合本人所开设的新课程《结构方程模型》(一学期三学分)进行教材的准备。尽管本书的章节数目与介绍内容分量繁多,却也只能对 SEM 进行大体概略式的介绍,范例举隅也是从 SEM 的诸多应用中选择较为具有代表性的范式进行说明。有一些非常值得介绍的应用范式(例如 MTMM 分析、成长曲线分析、交互作用分析),由于碍于篇幅,因此不在本书搜集的范例之内,有兴趣的读者可以从国外的文献中获得这方面的信息。

在工具方面,虽然本人撰写博士论文期间,是以 EQS 为分析工具,但是在本书的示范与说明中,本书仍选择了最传统也是最普遍被学术界接受的 LISREL 语言,希望能够将此书的概念与传统的 SEM 使用方式相结合。一旦读者对于 SEM 的基本原理有一定的认识与了解后,对于其他 SEM 软件应可以很快熟习。尤其是 Mplus 与 AMOS 两套软件,是非常值得推荐的 SEM 分析工具。Mplus 的优点是可以整合各种重要的统计分析技术,例如试题反应理论(IRT)、潜在类别模型(LCM)、多层次模型(MLM)。而 AMOS 则可以与 SPSS 软件搭配,使得高低阶的统计分析可以方便结合、巧妙搭配,适合硕博士研究生的使用。

总之,本书作为 SEM 的概说介绍与范例演示的教材,最主要的企图是推广 SEM 的使用,并使读者具有实践的能力。至于本书未能达到的目标与未能详细说明的内容,则有赖学术界的其他同好一起努力,使得 SEM 可以在国内的学术界占有一席之地,真正发挥其价值,为学术尽一些心力,是所至盼。

其後經多年發展，已成為全省重要之工業中心，其產物如羊毛、呢絨、棉織品等，銷路甚廣，且其交通、商業、文化、教育、衛生等，均極發達，為全省之冠。其地之重要，不僅在於其產物之豐富，且在於其交通之便利，其地之重要，不僅在於其產物之豐富，且在於其交通之便利。

其地之重要，不僅在於其產物之豐富，且在於其交通之便利。其地之重要，不僅在於其產物之豐富，且在於其交通之便利。其地之重要，不僅在於其產物之豐富，且在於其交通之便利。其地之重要，不僅在於其產物之豐富，且在於其交通之便利。其地之重要，不僅在於其產物之豐富，且在於其交通之便利。

其地之重要，不僅在於其產物之豐富，且在於其交通之便利。其地之重要，不僅在於其產物之豐富，且在於其交通之便利。其地之重要，不僅在於其產物之豐富，且在於其交通之便利。其地之重要，不僅在於其產物之豐富，且在於其交通之便利。其地之重要，不僅在於其產物之豐富，且在於其交通之便利。

其地之重要，不僅在於其產物之豐富，且在於其交通之便利。其地之重要，不僅在於其產物之豐富，且在於其交通之便利。其地之重要，不僅在於其產物之豐富，且在於其交通之便利。其地之重要，不僅在於其產物之豐富，且在於其交通之便利。其地之重要，不僅在於其產物之豐富，且在於其交通之便利。

附录一 LISREL 语法

LISREL 语法通常由三个部分所组成：数据设定、模型设定与输出设定。用户可以在打开 LISREL 软件后点选 **FILE** 当中的 **NEW** 的 **SYNTAX only** 开启一个新的语法窗口，以便输入 LISREL 语法，或以其他软件编写 LISREL 语法完成后读入。

LISREL 语法的长度，每一行上限为 127 位。原始数据或变量卷标等数据可以储存于外部档案，需要时再读入 LISREL 程序。外挂档案若以固定格式储存数据时，系统对于固定格式数据所支持的总长度为 1024 位。

LISREL 语法以 **TItal**（标题）为开端，**TI** 指令用来指定整个 LISREL 分析的标题名称，使每一页输出报表都有一个清楚的标题说明。如果标题放置于 LISREL 程序的首两行，可以省略主指令关键词 **TI**，LISREL 将自动视该一至二行文字为标题栏，直到遇见第一个主指令 **DA**。为了在 LISREL 语法当中加入一行无关于 LISREL 分析的标注说明时，可以用惊叹号 (!) 来进行批注，LISREL 将不会处理该行数据。有关 LISREL 其他各类指令与关键词的意义与使用方法说明如下。

一、数据设定指令

■ DAta

DAta 主指令用来定义数据。使 LISREL 可以理解研究者所输入的数据类型与内容。基本语法格式如下：

```
DA  NG=n  NI=k  NO=n  MA=type of matrix  
      XM=global missing value  RP=number of repetitions
```

NG number of groups

表示用于分析的样本组数，原始设定为 1。当研究者想比较不同的样本之间的模型估计的异同时使用。例如比较男生样本与女生样本时，数据输入区分成两

个群体，NG=2。比较三个学院（文艺、理工、法商）学生样本时，数据输入区分成三个群体，NG=3。

NI number of input of variables

表示输入的观察变量数目，原始设定为 0。变量数目系单指输入的观察变量数目，不包括非观察变量。当一个 LISREL 模型中包含 18 个观察变量时，NI=18。

NO number of observations

表示样本规模，即样本人数。当一个研究的数据库有 450 个受测者，NO=450。

XM a missing value label

用于遗漏值设定。当所使用的数据为原始数据时，可指定遗漏值。

MA the type of matrix to be analyzed

表示用于 LISREL 分析的矩阵类型（注意：此时所指的是 LISREL 进行估计的数据形态，而非指数据读入的形态）。可为下列几种选择：

MM matrix of moments about 0: 数值为 0 矩阵

CM a covariance matrix (default): 共变量矩阵（原始设定值）

KM a correlation matrix: 相关矩阵

AM an augmented moment matrix: 外增动差矩阵

OM a correlation matrix of optimal scores produced by PRELIS:
由 PRELIS 所产生的期望分数相关矩阵

PM a matrix of polychromic or polyserial correlations:
多元系列相关矩阵

RP number of repetitions

表示重复执行 LISREL 的次数，原始设定值为 1。用于 Monte Carlo 与 Bootstrap 研究。例如，当研究者希望重复执行 LISREL 分析 100 次，RP=100。

Data 指令的综合范例列举如下：

DA NI=10 NO=100 MA=CM

表示 LISREL 将读入 10 个变量 (NI=10)、100 个受试者 (NO=100) 的数据库, 数据分析形态为协方差矩阵 (MA=CM)

■ LAbels

LA bels 主指令用来进行输入变量的标签批注, 使 LISREL 输出结果与报表在呈现观察变量时, 可以易于理解的文字呈现。每一个卷标以八个位为限, 超过的位数则被忽略。卷标指令以 (‘ ’) 引号来描述标注内容, 且依观察变量的顺序排列, 指令格式如下:

```
LA FI=filename FO RE
format statement
'labels'
```

FI filename for containing the labels

存放观察变量标签批注的文件名。可以简化为 LA=filename。FO 与 RE 关键词则只配合外挂标签档案之使用, FO 表示卷标文字为固定格式时的格式形态。RE 表示标签档案读取后进行 rewound。当变量标签放置于外部档案时使用。

format statement

变量标签格式为固定格式时, 需使用 FORTRAN 语言的字符串格式来表示, 即 *A*, 第一个*代表变量数目, 第二个*代表位数 (最大值为 8)。如果要使用自由格式, 则在 LA 次行以*号表示即可, 标签之间以空白、逗号或换行表示区隔。

LA bels 指令范例列举如下:

1. LA

*

```
'HAPPY1' 'HAPPY2' 'HAPPY3' 'SES1' 'SES2' 'SES3' /
```

表示以自由格式界定标签, 共有六个观察变量的标签。最后一个斜线代表观察变量标签完毕, 如果此行所列出的变量标签个数等于观察变量数目, 斜线可以省略。

如果在任两个变量之间的某个变量没有标签, 可以将每一个标签的间隔改以逗号区隔, 再将没有标签的观察变量以两个逗号来代表跳过, LISREL 软件将以 VAR1, VAR2……来代替。例如想取消第二个变量卷标, 语法为:

```
'HAPPY1',,, 'HAPPY3' 'SES1' 'SES2' 'SES3' /
```

LISREL 将读到的标签为:

```
HAPPY1 VAR2 HAPPY3 SES1 SES2 SES3
```

2. LA

```
(6A7)
```

```
HAPPY1 HAPPY2 HAPPY3 SES1 SES2 SES3
```

表示观察变量标签将以固定格式表示。6A7 代表 LISREL 将以固定格式读取六个卷标, 每一个卷标“必须”占据七个字段(位)。

3. LA FI=label.txt

表示观察变量的标签存放于 label.txt 档案当中, 卷标的格式将在卷标档案当中的第一行中指定。如果是自由格式, 档案中的第一行可以直接放置变量标签, 或是标示一个*, 次行再出现标签。如果变量名称之后紧接着 FO 关键词, 那么在下一行应出现标签固定格式的界定值(*A*)。

■ Raw data

RAW data 主指令是用于当输入的数据为原始数据而非摘要统计数据(例如矩阵数据)时。而通常原始数据都储存在另一个档案中, 因此 RA 指令多跟随着路径关键词 FI。格式说明如下:

```
RA FI=filename FO RE
Variable format statement
data
```

原始数据如果放置于外部档案时，需要指定 FI 关键词并说明变量名称与路径。FO 关键词表示外部数据文件使用固定格式，因此在紧接下一行将会出现一个 FORTRAN 语言的数值格式 (*F*.) 符号，第一个*表示变量数目，第二个*表示每个变量所占据的字段数，第三个*表示小数点之后的位数，例如 5F6.3 代表有五个观察变量，每一个变量占六个字段，小数点三位。如果使用自由格式，则不需指定变量数据的格式，直接第一行便输入数据，或以*放置于第一行，第二行开始输入数据。

数据读取时，LISREL 有一些特殊的符号使用原则：第一，LISREL 将以数据间的空白（或是逗号）视为变量数据的分界点，依序自动读取每一个变量的数值。第二，当资料最后跟随一个斜线时，表示剩下的变量数值与前一个样本的剩下来的变量数值相同。第三，两个逗号之间的变量数值与前一个样本的该变量数值相同。第四，两个变量有相同的数值时，可以 n*data 来表示重复该数据，n 表示重复次数，data 表示数值。

RA 指令范例列举如下：

1. DA NI=5 NO=20

```
RA FI='c:\data.dat' FO
(5F3.1)
```

上述指令表示共有五个观察变量，20 个样本。数据形式为原始数据，放置于 C:\data.dat 路径与档案中，采取固定格式，共有五个观察变量，每一个变量占据三个位，其中一个为小数点。

2. DA NI=5 NO=20

```
RA
10.5 3.2 5.4, 5.8, 12.1
0.9 6.1 /
3.1 ,, 7.1 9.8 10
2*3.5 4.4
3.9 6.6 .....
```

此一范例表示有五个观察变量，20 个样本，数据内含于 LISREL 语法当中。每一个变量的数据包含了一个小数点，第二行的“/”代表第二个样本的剩余三个变量数据与

第一个样本的数据相同，两个逗号之间应补入上一个样本的相同变量的数值，“*”表示有两个变量数值都是 3.5。这些特殊符号使用的读入结果与下列的固定格式读出结果相同：

```
(5F3.1)
105 32 54 58121
    9 61 54 58121
31 61 71 98100
35 35 44 39 66
```

每一个变量数值占有三个字段（位），其中最后一个数字为小数点，因此 25 将被读为 2.5，121 将被读为 12.1。

■ CM, KM, MM, OM, PM

除了原始资料之外，LISREL 可以读入摘要统计数据，例如协方差、相关系数等。使用的指令为 **CM, KM, MM, OM, PM**。例如：

```
KM SY
1.00
.50 1.00
.40 .30 1.00
```

表示相关矩阵的数据（KM），且数据以三角矩阵方法读入对角线及以下三角形内的资料（SY）。如果矩阵数据储存于另一个档案中，可以以（FI）关键词指定文件名读取之。指令标准格式如下：

```
CM FI=filename SY FO RE
Variable format statement
Summary data statistics
```

上式中，CM 系指协方差矩阵，此指令的其他选择包括 KM（相关矩阵）、MM（动差矩阵）、OM（期望分数矩阵）、PM（多元系列相关矩阵）、RM（Spearman 等级相关

矩阵)、TM (Kendall's tau-c 相关矩阵) 等。

矩阵内部形态关键词包括了 SY 与 FU, SY 代表读取对角线及以下数据 (下三角区域), FU 代表读入矩阵中的所有数据。如果没有指定 SY 与 FU 任一关键词, 则矩阵数据应以固定格式形态, 依 SY 对角与下三角形的顺序, 读入所有的数据。

值得注意的是, 如果在 DAta 指令中, 数据分析形态指定为协方差矩阵 (MA=CM), 而数据读取形态为相关矩阵 (本指令使用 KM 关键词时), 还需输入各观察变量的标准偏差, 才能得到协方差的数据。另外, 如果在 DAta 指令中, 数据分析形态指定为零动时矩阵 (MA=MM), 而数据读取形态为协方差矩阵 (本指令使用 CM 关键词时), 还需输入各观察变量的平均数, 若读入为相关矩阵时, 各观察变量的平均数与标准偏差均需输入。平均数与标准偏差的读入请参考 ME 与 SD 指令之使用说明。以下列协方差矩阵 (S) 为例,

$$S = \begin{bmatrix} 20.78 & -5.54 & 10.21 \\ -5.54 & 25.90 & 7.67 \\ 10.21 & 7.67 & 21.25 \end{bmatrix}$$

以固定形态输入时, 矩阵指令可为下列三种中任一种:

CM FU

(3F6.2)

```
2078 -554 1021
-554 2590 767
1021 767 2125
```

CM SY

(3F6.2)

```
2078
-554 2590
1021 767 2125
```

CM

(6F6.2)

```
2078 -554 2590 1021 767 1021
```

以自由格式输入时，语法如下：

CM FU

20.78 -5.54 10.21

-5.54 25.9 7.67

10.21 7.67 21.25

CM SY

20.78

-5.54 25.9

10.21 7.67 21.25

CM

20.78 -5.54 25.9 7.67 10.21

■ ME、SD

ME、SD 主指令用于读取观察变量的平均数 (ME) 与标准差 (SD) 数据。配合关键词可以读取放在外部档案的平均数与标准差资料。基本语法如下：

ME (or SD) FI=filename FO RE

Variable format statement

data (vector of NI means or standard deviation)

现以三个观察变量的平均数与标准差的读取为例，当以自由格式读入时如下：

ME

52.11 60.59 53.01

SD

4.64 4.12 5.88

平均数与标准差的读取可以与矩阵整合在一起，将平均数与标准差放置于矩阵数据之下，现以外挂数据与内含数据格式各举例如下：

```
DA NI=3 MA=MM
CM FI=DATA.COV
ME FI=MEAN.MEN FO
(3F6.2)
```

表示协方差矩阵数据储存于 DATA.COV 档案中, 平均数数据储存于 MEAN.MEN 档案中。

```
DA NI=3 MA=MM
CM ME FO
(3F6.2)
2078
-554 2590
1021 767 2125
5211 6059 5301
```

上例当中的最后一行为平均数数据。

■ Select

Select 主指令用于选取矩阵中的变量, 或将变量顺序重新排列。例如一个矩阵中有五个变量, 依 1、2、3、4、5 顺序排列, 或直接列出变量的名称。一般而言, LISREL 语法将 Y 变量放置于 X 变量之前。举例如下:

```
SE 5 4 3 2 1
```

表示将变量顺序倒置

```
SE 5 4 2 1
```

表示将变量顺序倒置且不选择第三个变量纳入分析。

二、模型界定指令

■ MModel

模型设定指令由 **MModel** 指令主导, 接统一连串的关键词来说明模型当中变量数目、类型与各种参数关系的界定。基本语法形态如下:

```
MO  NY=p NX=q NE=m NK=n AP=k FI
    LY=mf,mm LX=mf,mm BE=mf,mm GA=mf,mm
    PH=mf,mm PS=mf,mm TE=mf,mm TD=mf,mm
    TY=mf,mm TX=mf,mm AL=mf,mm KA=mf,mm
```

首先, MModel 指令必须说明模型中各重要变量的数目, 各关键词说明如下:

NY number of observed endogenous variables

内生观察变量(y)数目

NX number of observed exogenous variables

外源观察变量(x)数目

NE number of latent endogenous variables

内生潜在变量(η)数目

NK number of latent exogenous variables

外源潜在变量(ξ)数目

这四个主要次指令说明模型中各类变量的数目。LISREL 第八版新增加一个 **AP** (number of additional parameters), 纳入外加参数。一般传统 LISREL 模型多仅使用上述四种数目之说明。当一个模型同时用到这四种类变量, 表示为一 SEM 统合模型分析, 但也可能只用到其中几类, 例如, 当 NK 与 NX 合并使用时, 反映了外源变量的测量模型 (因素分析模型), 涉及了 LISREL 的 LX、PH、TD 矩阵的估计; 当 NE 与 NY 合并使用时, 反映了内生变量的测量模型 (因素分析模型), 涉及了 LISREL

的 LY、PS、TE 矩阵的估计；当 NX 与 NY 合并使用时，反映了观察变量之间的路径分析模型而无任何的潜在变量涉入其中，此时涉及了 LISREL 的 BE、GA、PS 矩阵的估计。

在 **Model** 指令中，如果跟随一个关键词 **FI**，LISREL 分析会自动将整个模型视为一个无潜在变量的模型。也就是产生下列设定：

$$\mathbf{NK}=\mathbf{NX}, \mathbf{\Lambda}_X=\mathbf{I}, \mathbf{\Theta}_\delta=\mathbf{0}, \mathbf{\Phi}=\mathbf{S}_{XX}(\text{fixed})$$

也就是将 **NK** 视为 0，没有潜在外源变量。因此，当 **FI** 出现在 **Model** 指令中时，表示 $\mathbf{\Phi}$ 矩阵（自变量间相关矩阵）将自由估计，并计算标准误 $\hat{\mathbf{\Phi}}$ 。当出现 **FI** 时，表示 $\mathbf{\Phi}=\mathbf{S}_{XX}$ 被设定为固定，且没有标准误 $\hat{\mathbf{\Phi}}$ 。

对于各矩阵 **LX**、**LY**、**BE**、**GA**、**PH**、**PS**、**TE**、**TD**、**TH**、**TY**、**TX**、**AL**、**KA**，**MO** 指令可以个别说明它们的矩阵格式 (matrix form; *mf*) 或形式 (matrix mode; *mm*)。如果矩阵的格式与形式同时被指定，*mm* 与 *mf* 则以逗号隔开（不得留空白）。各矩阵可能的形式与格式栏于表 6.2。以下列的 $\mathbf{\Phi}$ 矩阵为例：

$$\mathbf{\Phi} = \begin{bmatrix} 1 & & & \\ * & 1 & & \\ 0 & 0 & * & \\ 0 & 0 & 0 & * \end{bmatrix}$$

其中 * 表示自由估计参数，0 与 1 代表固定参数。可被写做：

```
MO ..... PH=FI .....
FR PH(2,1) PH(3,3) PH(4,4)
VA 1 PH(1,1) PH(2,2)
```

■ FR、FI、EQ

如果用户想要改变矩阵中的特定元素的关系，例如固定、自由、或限定参数的

估计时, 可以使用 FR、FI 与 EQ 指令。例如:

```
Free LX(1,2) LX(2,2)
```

表示让 LX 矩阵 (与 X 有关的因素负荷量矩阵) 中的第一与第二个元素的参数自由估计, 也就是第一与第二个观察变量与第二个因素的因素负荷量。

```
Fix LX(1,2) LX(2,2)
```

表示让 LX 矩阵 (与 X 有关的因素负荷量矩阵) 中的第一个与第二个元素的参数自由估计。

```
Equal LY(3,4) LY(4,4) BE(2,1) GA(4,6)
```

表示限定 LY 矩阵 (与 Y 有关的因素负荷量矩阵) 中的四个元素的参数具有相等的数值, 也就是第三与第四个观察变量与第四个潜在变量的两个因素负荷量, 第二个残差与第一个观察变量的结构参数以及第四个观察变量与第六个因素的结构参数等四个参数限定等同。

值得注意的是, 语法中的左右括号与逗号可以省略。例如下列两者是相同的设定:

```
FI LX(1,2) LX(2,2)
```

```
FI LX 1 2 LX 2 2
```

如果参数数目众多且具有连续的顺序, 可以减号 (-) 简化参数的指定, 例如下列两者是相同的设定:

```
FI GA(2,2) GA(2,3) GA(2,4) GA(2,5)
```

```
FI GA(2,2) -GA(2,5)
```

另外, 如果在 MO 指令中, 矩阵设定了 ZE 与 ID 限制时, 矩阵中的数据将不会

储存于计算机的内存中；若设定了 DI 限制时，只有对角线的数据被计算机读取，若设定了 PH=ST 限制时， Φ 矩阵的参数将无法在此地被设定。

■ 恒等性参数限定语法

MO 指令中，亦可以利用多样本分析对于参数进行跨样本的恒等限定。其做法是在第一个样本中，以 FR 指令设定该参数为自由估计，然后在第二个或所有后续样本中以 EQ 指令进行恒等限定。另一种做法是在 Model 指定中，在所需设定的矩阵（例如 LX、LY、PS 等）之后加入下列副指令：

- SP (same pattern)：表示样本间矩阵具有相同的形态。
- SS (same starting value)：表示参数估计使用与前一个样本相同的起始值来进行估计。
- PS (same pattern and starting value)：表示样本间矩阵具有相同的形态并使用与前一个样本相同的起始值来进行估计。
- IN (invariant)：表示矩阵为恒等矩阵，矩阵的形态与参数的估计状态（FR 或 FI），以及参数数值均与第一个样本相同。

有关跨样本的恒等设限与参数估计于第十一章与范例七当中予以说明。

■ CO、IR、PA、MA

CO、IR、PA、MA 是 LISREL 中特殊的参数限定与数据读入指令。比如说，使用者如果想要估计某一个参数限定为另一组参数的函数关系，可以用 COmplex equality constraints 指令来完成。此时，参数函数关系的表现可以利用一般数学四则运算符号+、-、*、/、**k（k 代表幂次）来连结各参数。范例如下：

```
CO TD(1,1)=1-LX(2,1)**2-LX(2,2)**2
CO LX(2,2)=LX(1,1)-LX(4,1)**4
CO BE(1,3)=-BE(3,1)
```

如果要将某一些参数的数值限定在一个特定区间当中或某个数值的上下之时，可使用 IR (interval restrictions) 指令进行：

```

IR TD(1,1)>0
IR GA(1,2)<1
IR PH(1,3)>-1 <1

```

值得注意的是，此时参数数值的范围虽然是以<与>符号来界定，但是实则是≤与≥的关系，也就是说，TD(1,1)参数的范围被限定在大于且等于 0；GA(1,2) 参数的范围被限定在小于且等于 1；PH(1,3)参数的范围被限定在小于等于 1 与大于等于-1。

参数限定除了以条列式的数值来陈述之外，也可以 0（表示固定参数）与 1（表示自由估计参数）的数值矩阵的方式来设定，此时需配合 **P**Attern matrix 指令进行。以Γ矩阵中的参数设定为例：

$$\Gamma = \begin{bmatrix} \text{free} & \text{fixed} & \text{fixed} \\ \text{fixed} & \text{free} & \text{free} \end{bmatrix}$$

对于上述Γ矩阵，可以利用四种不同的 **PA** 指令陈述法来表现：

```

PA GA
(6I1)
100011

PA GA
(3I1)
100
011

PA GA
*
1 0 0 0 1 1

PA GA
*
1 0 0
0 1 1

```

如果一个矩阵当中只有对角线必须设定, PA指令则只需对于对角线的参数进行指定。例如一个 Ψ (4×4)矩阵, 其中的 ψ_{11} 与 ψ_{33} 为固定参数, ψ_{22} 与 ψ_{44} 为自由估计参数, 则可以PA指令表示为:

PA PS

*

0101

如果以矩阵输入的不是 0 与 1 的形态设定值, 而是具体数值, 此时应使用 **MA**trix **values** 指令来设定, 例如下列 **B** 矩阵的 **MA** 表示法:

$$B = \begin{bmatrix} 0 & 0 & 0 \\ .5 & 0 & 0 \\ .5 & .5 & 0 \end{bmatrix}$$

MA BE

(3F1.1)

000

500

550

MA BE

*

0 0 0

.5 0 0

.5 .5 0

■ VA、ST

Value 与 **S**Tarting value 两个次指令用来说明当参数限定为 0 之外的数值, 也就是估计的起始值。这两个次指令有相同的数学性质, 因此可以互换使用。

VA 1.5 LX(2,1) LX(6,2) GA(1,2)

表示令LX矩阵中的两个元素与GA矩阵中的一个元素参数值为 1.5。

■ LE、LK

LE 与 LK 两个次指令用来给予潜在变量文字卷标, LE 为潜在内生变量标签, LK 为潜在外源变量卷标, 两者功能与标示原则与 LA 指令相同。基本语法如下:

```
LE(or LK) FI=filename FO RE
format statement
'labels'
```

较简单的做法是在 LE 与 LK 指令之后加上一个分号 (;) 后, 直接加入各潜在变量的标签。例如:

LE; 幸福感 自尊

三、结果输出设定

LISREL 软件可以提供多种不同的参数估计方法与输出信息, 除了原始设定的一些估计方法与统计报表之外, 还有多种可供选择的报表。LISREL 出厂设定为使用最大概率法 ML 或 UL 法进行参数估计, 固定输出报表为模型的契合度指数与描述统计资料。参数估计与结果输出指令的下达以 OUtput 主指令接续适当的次指令或关键词为之。基本语法如下:

```
OU ME=methods RC=c SL=100 $\alpha$  NS RO AM SO
```

■ MEmethod (method of estimation) 参数估计方法

输出设定当中, 使用者必须设定参数估计的方法, LISREL 提供了多种不同的估计方法, 关键词说明如下:

- IV** instrumental variable method
- TS** two-stage least squares
- UL** unweighted least squares
- GL** generalized least squares
- ML** maximum likelihood (default)
- WL** generally weighted least squares
- DL** diagonally weighted least squares

除了上述进行参数估计的方法选择之外，对于输出设定的其他关键词尚包括下列各种状况：

- RC** ridge constant 脊状常数处理
矩阵数值重复乘以 10，直到矩阵可以为正常定位。原始设定值 RC=.001。
- SL** significance of modification procedures
修正程序的显著性，以百分比数值表示，例如 SL=5 表示 .05 的显著水平。原始设定值 SL=1。
- NS** do not compute starting value
计算机不执行起始值估计，而需由用户利用 ST 或 MA 指令来指定起始值。
- RO** ridge option (used with RC)
当协方差矩阵无法正常定位时，自动执行脊状常数处理。
- AM** automatic model modification
LISREL 将自动进行模型修正，方法是逐次将最大的未估计参数改设定为估计参数并加以估计。
- SO** scaling check off
关闭 LISREL 中自动对于潜在变量进行尺度化的功能。

■ 报表输出设定

LISREL 的报表中除提供数据读取之后的各种描述统计资料外，尚提供多种不同的统计数据，由用户自由选择，关键词如下：

- SE** standard errors
标准误
- TV** t values
报告 t 检验值
- PC** correlations of estimates
列出参数估计的相关系数
- RS** residuals, normalized residuals, and Q plots
列出残差、正态化残差与 Q 图
- EF** total effects and indirect effect
列出总效应与间接效应
- MR** miscellaneous results (=RS, EF)
列出 RS、EF 等多重数据
- MI** modification indices
列出模型修正指针
- XM** suppress computation of the modification indices
列出精简的修正指标
- XI** limit the fit statistics
仅列出一行模型契合度数据（卡方、自由度与显著性）
- FS** factor scores regressions
列出因素分数回归值
- SS** standard solution
标准化参数估计解
- SC** solution completely standardized
参数估计解完全标准化
- AL** print everything
列出所有 LISREL 分析的统计报表
- TO** print 80 characters/line
正常打印，每行打印 80 字符。为 PC 设定值

WP print 132 characters/line

宽式打印，每行打印 132 字符

ND number of decimals

报表中的小数点数目，可以为 0-8 任意数。原始设定值 ND=2

■ 估计迭代设定

LISREL 软件进行参数估计时，迭代程序可以利用下列方式来操作：

OU **TM**=*t* **IT**=*n* **AD**=*m* **EP**= ϵ

TM maximum number of CPU seconds allowed for current problem

微处理器最大处理秒数，例如 TM=60，表示微处理器进行迭代最大时间为 60 秒。原始设定值为 TM=172800（二天）。

IT maximum number of iterations

最大迭代次数，原始设定值为参数估计数的三倍。

AD check the admissibility of the solution

估计解接受度检测，当进行 *m* 次迭代后进行，若接受度不足则迭代停止，例如 AD=10，表示 10 次迭代后即检查估计解接受度。

EP convergence criterion(epsilon)

收敛值标准指定，原始设定值 EP=.000001。

■ 矩阵输出设定

LISREL 软件可以将特定矩阵输出至外部档案以供检视，基本语法如下：

OU *matrix1=filename1 matrix2=filename2*

LISREL 分析得到的矩阵列于左侧，文件名与路径置于右侧。文件名由使用者任意指定之。而矩阵可为 LY, LX, BE, GA, PH, PS, TE, TD, TY, TX, AL, KA。此外，LISREL 尚提供下列各种档案输出设定选项，且以固定格式（6F13.6）列于档案当中：

- MA** reordered input matrix
重新排序后的输入矩阵
- SI** the fitted matrix, $\hat{\Sigma}$ (Sigma matrix)
导出协方差矩阵
- RM** regression matrix of latent on observed variables
基于因素分数对观察变量的回归系数矩阵
- EX** estimated asymptotic covariance matrix of LISREL estimates
估计渐进协方差矩阵
- GF** all the goodness-of-fit measures
所有模型拟合评估值
- PV** the vector of estimated free parameters
自由估计参数向量
- SV** the vector of corresponding standard errors
标准误差向量
- TV** the vector of corresponding t-values
t 检验值向量

■ 路径图输出设定

如果使用者想要得到 LISREL 分析完成之后包含有各种参数估计结果的路径图, 可以用 **PD** (path diagram) 指令, 要求 LISREL 绘制模型路径图。**PD** 指令需放置于 **MODEL** 指令之后。

LISREL 一旦执行完毕之后, 会将所有的数据以路径图的方式加以整理, 使用者可以打开结果类型对话卷帘, 选择所需要的数据类型。LISREL8 路径图输出共提供了参数估计、标准化解、概念图、t 检验值、修正指标、预期改变量等六种路径图结果输出模式。

此外, 利用 **档案** (FILE) 当中的 **Export As Metafile**, 则可将路径图连同分析数据转换成图形文件, 轻易地与其他软件 (例如 Word、PowerPoint) 结合使用。

附录二 SIMPLIS 语法

一、简介

SIMPLIS 语法可以在第八版之后的 LISREL 软件中来执行。LISREL8 同时接受两种程序语言的命令，LISREL 语法已在附录一中加以介绍，SIMPLIS 语法在本附录加以介绍。为了区分这两种语言，以 LISREL 语法所撰写的语法程序称为 LISREL 输入 (LISREL input)，以 SIMPLIS 所撰写的语法程序称为 SIMPLIS 语法输入 (SIMPLIS input) (Jöreskog & Sörbom, 1993)，两种语法的撰写方式不同，不可混用。

SIMPLIS 分析所得到的输出报表有两种形式：SIMPLIS 格式或 LISREL 格式。一个 SIMPLIS 输入经过执行之后，可以产生 SIMPLIS 语法专属的报表格式，也可以得到与 LISREL 分析相仿的输出报表。SIMPLIS 格式的报表与 LISREL 格式最大的不同是采用方程式的输出模式，将传统 LISREL 报表中巨细靡遗出现的各种参数或检验值，直接以容易理解的方程式来整理呈现。相对的，如果产生的是 LISREL 格式报表，用户必须从报表当中找寻有意义的数据，对于不熟悉 LISREL 分析原理的用户来说，是相当困难的一件事。

虽然 LISREL 格式报表难以阅读、需要进一步的统整，但是却是被 SIMPLIS 软件的作者建议采用。因为首先 LISREL 报表提供的是参数矩阵的数据，符合 SEM 分析的基本概念；其次，LISREL 报表较 SIMPLIS 格式报表详细。所以，SIMPLIS 的作者直接将 SIMPLIS 语法与 LISREL 报表相链接，使操作者可以自由选择以 SIMPLIS 报表还是 LISREL 报表来呈现数据，所需要的指令只有简单的一行：

LISREL Output

或更具体地指明需要 LISREL 报表的哪些项目，例如：

LISREL Output: RS MI SS SC EF

冒号后面的选项，其意义与功能完全与 LISREL 分析相同。基于 SIMPLIS 可以获得与 LISREL 分析一样的报表，本书将仅介绍 SIMPLIS 的语法，范例中的报表仍以

LISREL 格式报表为主。对于 SIMPLIS 报表有兴趣的读者可以参考 SIMPLIS 操作手册的介绍。

二、SIMPLIS 语法的基本内容

SIMPLIS 语法的基本概念与 LISREL 相似，是以 ASCII 字符来撰写，大写或小写均可，但是变量名称必须为大写，如果要给变量加以特殊标签，可以利用左右上单逗点（‘ ’）来进行标签，单逗号内可以输入中文字形。语法以 TITLE 指令作为开端，用以说明整个 SIMPLIS 分析的标题，紧接着包括几个部分。第一，数据的定义，主要在说明数据的格式与内容，包括观察变量的名称与卷标、输入数据的形态、样本大小。第二，模型界定指令，说明研究者所欲检验的 SEM 假设模型，包括潜在变量的名称与标签、变量的关系、路径的类型与内容、潜在变量的量尺化、起始值设定、误差项的处理、限制参数的处理。第三部分是有关报表输出的设定，包括打印的宽度、小数点的位数、参数估计的方法、估计迭代次数设定与 LISREL 报表的指定等。

SIMPLIS 语法可以惊叹号“!”或“/*”符号说明指令开始的位置，或直接开始语法的撰写。指令的终点可以直接换行，或以分号“;”表示指令的结束。同一行中，可用分号连接多个指令以节省空间。

SIMPLIS 语法的第一行为标题栏，标题栏的行数不限，也可以不输入任何标题。通常使用者会利用前几行的空间，自由输入与该 SIMPLIS 分析有关的文字，作为辨识该语法的内容之用。其中第一行文字则被视为该 SIMPLIS 报表的标题栏，将出现在每一页报表的最开端（在标题栏与变量标签中，皆可输入中文字形）。

在标题栏之后，SIMPLIS 会以 Observed variables、labels 或 DA、Da、dA、da 当作是指令的开始。而 SIMPLIS 指令的下达是以一段与英文文法相通的文字作为开端，例如：

Covariance matrix from file EX1.COV

Sample size = 254

上述指令指出协方差矩阵数据放置于 EX1.COV 这一个档案中，样本数为 254。此一特色使得 SIMPLIS 语法易读易懂，语法的内容遵循一般的英文文法，阅读 SIMPLIS 语法与阅读英文语句一样的方便。在数学方程式的表现上，亦与一般数学的使用惯例相仿，使用者仅需具备高中程度的数学知识，即可顺利撰写 SIMPLIS 语法。在 SIMPLIS 语法的终了最好以指令：

End of problem

作为终结，将之视为语法程序终结的标示。以下，仅就 SIMPLIS 语法的内容加以介绍。

(一) 数据界定指令

■ Observed Variables 与 Labels

SIMPLIS 语法的第一个指令是用来说明观察变量的名称与卷标，语法为 Observed Variables 或是 Labels（此两个指令是相同的作用）。例如：

Observed variables: PERFORM MOT1 MOT2 ABLE1 ABLE2

或

Observed variables: '绩效' '动机 1' '动机 2' '能力 1' '能力 2'

在 Observed variables 或 Labels 指令之后，空一格，即可开始输入变量名称，或是直接输入变量标签（不必另行输入变量名称，整个单逗号内的文字连同单逗号都是变量名称）。如果用户想要获得中文标签，必须在单逗号内输入中文字，以整组卷标连同单逗号作为变量名称。

变量的名称或卷标数据也可以储存在外部档案，再读入 SIMPLIS 当中：

Observed variables from file= *filename*

或

Labels from file= *filename*

当变量数目众多时，可以利用连续数字作为变量名称，以短横线来说明变量的起讫，例如两百个变量，可以命名为 VAR1 到 VAR200，或 A1 到 A150 与 B1 到 B50，再以下列语法输入：

Observed variables: VAR1 – VAR200

Observed variables: A1 – A150 B1 – B50

■ 数据格式指令

SIMPLIS 中, 数据格式指令的运用与 LISREL 语法的 DATA 指令功能类似, 用以说明数据的格式。包括 Raw data、Covariance matrix、Covariance matrix and means、Correlation matrix、Correlation matrix and standard deviations、Correlation matrix, standard deviation, and means 几种形态。

当以原始数据读入时, 最好以外部数据文件来储存原始数据, 再以自由格式读入, 语法如下:

Raw data from file = *filename*

值得注意的是, SIMPLIS 无法处理遗漏值, 因此若有遗漏值应以 PRELIS 语法处理。

协方差矩阵或相关矩阵可以直接在指令下方输入, 以三个观察变量的矩阵为例, 语法如下:

Covariance matrix

10.03

2.86 13.97

5.01 4.59 19.80

Covariance matrix

10.03 2.86 13.97 5.01,4.59,19.80

若辅以 FORTRAN 语言的数字格式, 矩阵中的数据可以省去小数点, 标以矩阵的向量数目 (6) 或变量数 (3)。如下所示:

Covariance matrix

(6F5.2)

1003 286 1397 501,459,1980

Covariance matrix

(3F5.2) SYMMETRIC

```

1003
286 1397
501 459 1980

```

数字间可以留空白或以逗号区隔之。指令完毕后亦可以冒号直接输入数字。以节省空间，如下所示：

```
Covariance matrix: 10.03 2.86 13.97 5.01,4.59,19.80
```

若欲输入平均数或标准差数据时，则可以下列方式处理：

```

Means: 40.08 2.03 12.20
Standard deviations: 0.125 0.098 1.25

```

■ 样本数

SIMPLIS 的样本数指令如下：

```
Sample size = number
```

或

```
Sample size: number
```

或

```
Sample size is number
```

(二) 模型界定指令

■ 潜在变量的指令

模型界定最主要的工作是指出潜在变量为何。指令如下：

```
Latent variables: variablename(s)
```

或

```
Unobserved variables: variablename(s)
```

如果没有潜在变量，此一指令可以省略。欲进行中文标签时，可以在上单逗点内直接输入中文字形：

Latent variables: '中文变量名称'

■ 变量关系指令

为了说明变量之间的关系, SIMPLIS 语法以方程式直接表明变量关系, 指令为 Relationships, 等号左边为目的变量, 等号右边为条件式, 说明对目的变量的影响条件, 例如:

Relationships:

$$A1 = 1 * F1$$

$$A2 = F1$$

$$A3 = F1$$

$$B1 = 1 * F2$$

$$B2 = F2$$

$$B3 = F2$$

$$F2 = F1$$

上述指令表示有两个潜在变量 F1 与 F2, 各有三个观察变量 A1 到 A3 与 B1 到 B3 来估计这两个潜在变量。其中两个潜在变量的第一个观察变量因素负荷量设定为 1。最后一行指出 F2 被 F1 预测, 因此, F2 为内生潜在变量, F1 为外源潜在变量, F2 与 F1 之间具有一个 γ 参数说明 F1 对 F2 的回归系数。

除了以上述的方程式形式来设定变量关系, SIMPLIS 可以利用路径指针的形式来指定变量关系, 指令为 Paths, 语法为从 *variables -> variables*:

Paths:

$$F1 \rightarrow 1 * A1 \quad A2 \quad A3$$

$$F2 \rightarrow 1 * B1 \quad B2 \quad B3$$

其中的 \rightarrow 表示直接效应。

如果 A3 观察变量除了被 F1 影响, 也被 F2 影响, 与 A3 有关的路径为:

$$A3 = F1 \quad F2$$

或

F1 F2 -> A3

■ 潜在变量的量尺化

为了使潜在变量具有适当的量尺, SEM 模型必须对潜在变量进行量尺化, 例如指出某一个因素负荷量为 1 的固定值。在 SIMPLIS 语法中, 仅需在关系指令中的适当参数直接给定所欲固定的数值。如前述的语法范例:

A1 = 1 * F1

A2 - A3 = F1

第一个观察变量的因素负荷量设定为 1, 以设定 F1 的量尺。

■ 残差项设定

SEM 模型中, 残差项的设定通常包括测量模型的测量残差 (error in x or y) 与结构模型的解释残差 (error in Eta) 两种类型。在 SIMPLIS 的语法中, 如果要将残差项的变异量设定为 1, 设定方法如下:

Let the error variances of *variablename(s)* be 1

或

Set the error variances of *variablename(s)* equal to 1

若要设定残差间具有相关, 做法如下:

Let the errors between *variablename1* and *variablename2* correlate

或

Set the error Covariance between *variablename1* and *variablename2* free

若要设定残差间的相关为 0, 做法如下:

Set the error Covariance between *variablename1* and *variablename2* to 0

■ 因素间相关

一般而言, 在没有特别指定的情况下, SEM 模型中的外源变量的相关是自由估

计的参数,但是如果要因素间的相关的估计取消,做法如下:

Set the Covariance of *factorname1-factorname2* to 0

■ 参数的相等性假设

有时,在 SEM 模型中,我们会将某些参数设定为相等,使两个参数视为同一个参数来估计。在 SIMPLIS 中,做法如下:

Set the path from *variablename1* to *variablename2* equal to the path from *variablename3* to *variablename4*

或

Set path from *variablename1* to *variablename2* = path from *variablename3* to *variablename4*

或

Set path *variablename1 -> variablename2* = path *variablename3 -> variablename4*

或

Set *variablename1 -> variablename2* = *variablename3 -> variablename4*

对于残差变异量设定为相等,做法如下:

Set the error variances of *variablename1* and *variablename2* equal

或

Let the error variances of *variablename1* and *variablename2* be equal

或直接以下列指令设定之:

Equal error variances: *variablename(s)*

三、输出设定指令

■ Options 指令

SIMPLIS 语法对于分析完成后的报表打印格式,可以用 Options 指令来进行下列

设定:

RS: Print residuals (列出残差)

WP: Wide print (宽式报表打印, 132bits 格式)

ND= n : Number of Decimals = n (小数点位数)

ME= *options* : Method of estimation = *options* (参数估计方法)

AD= *off/on* : Admissibility check = *off/on*

IT= n : Iterations = n (迭代次数)

SI=*filename*: Save sigma in file *filename* (结果储存于外部档案)

完整的语法如下:

Options: RS WP ND= n ME=*options* AD=*off/on* IT= n SI=*filename*

其中, 估计方法的选择有七种不同方式, 包括:

Instrumental variable (IV)

Two-stage least squares (TSLS)

Unweighted least squares (ULS)

General least squares (GLS)

Maximum likelihood (ML)

Generally weighted least squares (WLS)

Diagonally weighted least squares (DWLS)

■ LISREL Output 指令

SIMPLIS 语法的报表有两种不同的形态, 如果以前述方式来进行 SIMPLIS 估计, 将得到 SIMPLIS 的方程式形态的报表。利用下列指令则可以得到与 LISREL 分析相同形式的结果报表。

LISREL Output

附录三 LISREL 常见问题

一、语法错误与执行错误

最常见的 LISREL 技术错误是语法撰写上的疏忽所造成,也就是语法错误(syntax error)。在第六章我们已经详细介绍了 LISREL 的语法指令,但是仅能对于语法的撰写原理进行介绍,在真实的应用上,仍会有许多问题是无法在该章节中涵盖的。同时,产生语法撰写错误的原因有各种可能,例如少了一个逗号、关键指令的代号拼字错误、指令的顺序错误等等,必须由用户自行检查语法中错误发生的原因。本书将配合范例,详细介绍 LISREL 指令的撰写方式,读者可以自行参阅,并多加练习。这无疑将可以避免此类的错误。如果读者实在不能适应 LISREL 繁复的语法指令,可以改用 LISREL 所提供的 SIMPLIS 指令进行 LISREL 分析。

语法错误的结果通常会导致执行错误,LISREL 可以在读入语法的同时,即可早期发现这些基本执行错误,但是有些比较严重的错误,必须在执行过程中随着分析的进展才能侦测到错误。造成执行错误的常见原因包括:

1. 符号的缺漏。例如缺少该有的斜线、逗号、空格键等。
2. 错误的指令或拼字错误。例如完全标准化解的输出指令为 SC,被误作 CS。
3. 错误的指令顺序,例如 DA、MO、OU 等指令没有依照顺序出现。
4. 无效的命令。尤其容易发生在模型设定指令(MODEL),例如 NX 与 NK、NY 与 NE 没有搭配使用,或矩阵的定义没有依照 LISREL 的限制。
5. 参数指定无效。容易发生在 FR 指令与 FIX 指令,错把 Beta 参数作为 Gamma 参数。

这些语法错误所造成的技术错误称之为基本操作错误,通常会在 LISREL 读完语法指令之后,立即以 ERROR message 出现在报表中,同时停止一切分析工作,直到使用者找出错误为止。

二、严重执行错误

前面所提到的错误往往是 LISREL 无法进行分析，但是有些错误并非是语法错误所造成，而是由于数据本身的问题 (data errors) 或是模式设定的不当 (specification errors) 所造成。通常会在产生问题的分析阶段，出现错误讯息提醒研究者错误的存在。但是这些数据或模型的错误通常不会造成分析中断，因此分析的正确性堪虑。

为了简化 LISREL 语法的内容，SEM 分析所使用的数据最好是放在外部的档案，以减少因为数据本身输入错误或过于冗长而造成检查上的不方便。因为当 LISREL 语法过长之时，LISREL 软件检查错误的能力也就受到限制。

■ 资料矩阵的问题

数据上的错误多半与输入的数据库有关，例如下列的讯息表示矩阵的数据是不合理的：

F_A_T_A_L_E_R_R_O_R: Matrix to be analyzed is not positive definite.

当输入矩阵数据存在不合理数据，使得分析过程无法有效处理协方差矩阵时，会提醒使用者遇到了非正定问题。此时，使用者必须回头检查所使用的矩阵数据有何问题。例如，当变量的数值是经过人工的合并调整，将某几题合并为一个题目且题数不同时，会导致某些变量变异数特别大，某些题目特别小，从而造成变量共变量矩阵的非正定问题。解决方法是重新整并数据，将合并的新变量除以题数，使各变量具有相仿的方差，或许才能解决此一问题。

■ 初始估计数的问题

另一个能发生的错误，是 SEM 分析使用不适当的起始值 (starting value) 所造成。通常如果一个 SEM 模型关系过于复杂，或是变项关系过于微弱，LISREL 必须耗费较长的估计时间与较多的迭代次数来进行估计。由于数据内部的关系不佳，致使 LISREL 从矩阵资料导出的第一次估计数距离观察数据真实参数数值过远，

LISREL 产生下列讯息来提醒用户，并提出操作建议：

F_A_T_A_L E_R_R_O_R: Unable to start iterations because matrix SIGMA is not positive definite. Provide better starting value.

上述问题的解决，是利用 ST 指令来指出较为接近事实的起始值。使用者可以利用其他的途径来“预估”这些参数可能的数值，然后输入 LISREL 语法中，可以有效提升 LISREL 分析的效率。

■ 接受检查点错误

LISREL 的执行错误可能被接受检查点 (admissibility checkpoint) 所侦测出来。LISREL 分析在 20 次迭代之内，不会进行模型接受度检验，但是超过 20 次迭代仍未能获得理想的估计数之后，即会检查 SEM 模型产生最后终解的适当性。

接受度检验的内容是测量变量的因素负荷量 (Λ_x 、 Λ_y) 是否合理 (例如是否出现 0 的负荷量) 以及方差共变量矩阵 (Φ 、 ψ 、 Θ 矩阵) 是否具有非正定问题。如果检查的结果发现问题，将出现下列错误讯息：

F_A_T_A_L E_R_R_O_R: Admissibility check failed.

当上述讯息出现的同时，LISREL 也停止估计工作，并且在出现问题的矩阵下方，出现一个警告讯息 (warning message)，指出是哪一个参数出现了问题。对于使用者来说，接受度检验是一个非常简便的察觉模型问题的工具。

经过检视之后，如果用户认为数据与模型都是正确无误的，单纯的只是因为模型复杂或变量关系本身的限制，可以将进行接受检验的迭代次数调整到更高的水平，使用的指令为 AD 指令，例如 AD=30 表示 LISREL 分析要到第 30 次迭代才进行接受检验，或甚至关闭接受检查 (AD=OFF)。但是一般均不建议关闭 AD 功能，除非模型当中，对于某些矩阵 (Φ 、 ψ 、 Θ 矩阵) 进行 ZERO 矩阵的设定 (Jöreskog & Sörbom, 1989)。

■ 收敛问题

当 LISREL 分析无法在预设的迭代次数内完成参数估计时，会产生收敛失败问

题，并造成 LISREL 分析的中断。在 LISREL 中，迭代次数是以自由估计参数的三倍作为默认值。当此一问题发生时，LISREL 会产生下列讯息：

W_A_R_I_N_G: The number of iterations exceeded XX.

F_A_T_A_L E_R_R_O_R: Serious problems encountered during minization.
Unable to continue iteration. Check your model and data.

此一问题的发生主要是因为假设模型与真实数据之间差异过大，也就是模型设定的不适当，造成估计的失败。改善的方法除了调整迭代次数（IT 指令），就是回去检查模型的设定是否有问题。用户也可以利用 PC 指令，要求 LISREL 列出参数估计数的相关系数来检查是哪些参数之间具有非正定问题。透过这些检查，找出可能造成错误的参数。据以调整模型，再次进行 LISREL 分析，或许可以解决无法收敛的问题。

附录四 SEM 操作常见疏失检核表

Kline (1998) 曾经撰文指出一般使用者会犯下的 SEM 操作错误的一些技术细节，十分值得参考，列举如下：

一、与设定有关的疏失

1. 在数据收集完成之后才开始进行模型界定。
2. 忽略了结构模型当中潜在变量的关系。
3. 潜在变量缺乏足够的测量指标。
4. 测量尺度不恰当，不符合心理计量的要求。
5. 对于变量的因果关联缺乏谨慎的考虑。
6. 利用回溯关系的设定，以掩饰变量影响方向不确定。
7. 为了追求模型拟合度，不惜改变模型设定，并使模型变得非常复杂。
8. 没有充分的理由下，将测量残差或解释残差的相关纳入估计。
9. 没有充分的理由下使用多维假设（同一个测量变量受到一个以上的潜在变量的影响）。

二、与资料有关的疏失

10. 忽略了对于输入数据最基本的检查。
11. 忽略了遗漏值的遗漏形态（随机或系统化遗漏）的检验。
12. 忽略了数据的次数分配的检查。
13. 忽略了偏离值的检查。
14. 忽略了线性假设的检查。

三、与执行过程有关的疏失

15. 完全基于统计数字来判断模型的修饰。
16. 忽略了分析软件的语法指令的检查。
17. 不正确地使用相关矩阵作为输入数据。
18. 变量的相关过高，造成不稳定的结果。
19. 用小样本分析很复杂的模型。
20. 对于潜在变量量尺化的操作有误。
21. 忽视 SEM 的起始值所带来的影响，或设定了不恰当的起始值。
22. 没有使用适当或必要的方法去确认模型的辨识度是否足够。
23. 没有正确有效地避免实务上辨识不足的问题。
24. 对于统合模型，没有分别针对测量模型与结构模型进行模型评估。

四、与结果报告有关的疏失

25. 仅关心模型的整体契合度，而忽略了其他指标。
26. 将模型契合度解读为模型被支持的唯一或主要证据。
27. 将模型契合度解读为内生变量被有效解释的证据。
28. 过度倚赖显著性检验的数据与分析结果。
29. 以不适当的方法解读标准化参数数据的意义。
30. 对于完全相同的模型 (equivalent models) 没有妥善地处理。
31. 没有妥善地运用替代模型。
32. 将潜在变量 (因素) 以具体变量来解释。
33. 错认为 SEM 可以弥补一个设计不佳或思虑不周的研究。
34. 没有善尽一个研究者的责任，提供充分的信息与正确的解读给读者，使他人无法理解或复制研究结果。
35. 从显著的参数数据，过度乐观地下达因果结论。

参考文献

中文部分

- 王木荣, 林幸台. 威廉斯创造力测验修定研究. 台湾: 师大特殊教育研究学刊, 1986, 2, 231-250.
- 吴静吉, 高泉丰, 王敬仁以及丁兴祥. 拓弄思图形创造思考测验指导及研究手册(甲式). 台湾: 远流出版社, 1981.
- 吴静吉, 王文中, 郭俊贤, 陈淑惠, 李慧贤. “我的老师”问卷之修订. 未出版手稿, 1996.
- 林碧芳, 邱皓政. 中小学教师创意教学自我效能与相关因素之研究. 台湾: 辅仁大学主办“中国”心理学学会第四十二届年会, 2003.
- 邱皓政. 组织创新环境的概念建构与测量工具发展. 台湾: 政治大学主办创造与创新研讨会, 2003.
- 邱皓政. 青少年自我概念向度与成份双维理论之效度检验与相关因素研究. 台湾: 教育与心理研究, 1999, 26(1), 85-131.
- 郭美凰. 新进人员信息搜寻行为、组织社会化程度与组织行为结果之相关研究. 台湾: 云林技术学院企业管理技术研究所未发表硕士论文, 1996.
- 叶玉珠, 吴静吉, 郑英耀. 影响科技与信息产业人员创意发展的因素之量表编制. 台湾: 师大学报: 科学教育类, 2000, 45(2), 39-63.

英文部分

- Aitken, A. C. (1934). Note on selection from a multivariate normal population. *Proceedings of the Edinburgh Mathematical Society*, 4, 106-110.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317-322.
- Allison, P. D. (1987). Estimation of linear models with incomplete data. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 71-103). San Francisco: Jossey-Bass.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice-Hall.
- Anderson, J. C., & Gerbing, D. W. (1984). The effects of sampling error on convergence, improper solution and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49, 155-173.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411-423.
- Anderson, J. C., & Gerbing, D. W. (1992). Assumptions and comparative strengths of the two-step approach: Comment on Fornell and Yi. *Sociological Methods & Research*, 20 (1), 321-333.

- Bagozzi, R. P. (1983). Issues in the application of covariance structure analysis: A further comment. *Journal of Consumer Research*, 9, 449-450.
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74-94.
- Bagozzi, R. P., & Phillips, L. W. (1992). Representing and testing organizational theories: A holistic construal. *Administrative Science Quarterly*, 27(3), 459-489.
- Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling*, 1, 35-67.
- Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269-296). Mahwah, NJ: Lawrence Erlbaum.
- Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, 9, 78-102.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality & Social Psychology*, 51, 1173-1182.
- Barrett, P. T., & Kline, P. (1981). Radial parcel factor analysis. *Personality and Individual Differences*, 2, 311-318.
- Bearden, E. O., Sharma, S., & Teel, J. E. (1982). Sample size effects on chi-square and other statistics used in evaluating structural equations. *Journal of Marketing Research*, 19, 425-430.
- Bentler, P. M. (1988). *Theory and implementation of EQS: A structural equations program*. Newbury Park, CA: Sage.
- Bentler, P. M. (1989). *EQS Structural Equations Program Manual*. Los Angeles, CA: BMDP.
- Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological Bulletin*, 112, 400-404.
- Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Bentler, P. M., & Yuan, K-H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, 34, 181-197.
- Bobko, P., & Rieck, A. (1980). Large sample estimators for standard errors of functions of correlation coefficients. *Applied Psychological Measurement*, 4, 385-398.
- Bollen, K. A. (1989). *Structural equation modeling with latent variables*. New York: John Wiley.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605-634.
- Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, 107, 260-273.
- Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137-164). Newbury, CA: Sage.
- Browne, M. W., & Du Toit, S. H. C. (1992). Automated fitting of nonstandard models. *Multivariate Behavioral Research*, 27, 269-300.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Press.
- Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows*. Newbury Park, CA: Sage.
- Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.
- Byrne, B. M., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology*, 30, 555-574.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Carmines, E. G. & McIver, J. P. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G. W. Bohrnstedt and E. F. Borgotta (eds.), *Social Measurement: Current Issues* (pp. 65-115). Beverley Hills, CA: Sage.
- Cattell, R. B., & Baggaley, A. R. (1960). The salient variable index for factor matching. *British Journal of Statistical Psychology*, 13, 33-46.
- Cattell, R. B. (1974). Radial item parcel factoring vs. item factoring in defining personality structure in questionnaires: Theory and experimental checks. *Australian Journal of Psychology*, 26, 103-119.
- Cattell, R. B., & Bursdal, C. A. (1975). The radial parcel double factor design: A solution to the item-versus-parcel controversy. *Multivariate Behavioral Research*, 10, 165-179.
- Cattin, P. (1980). Note on the estimation of the squared cross-validated multiple correlation of a regression model. *Psychological Bulletin*, 87, 63-65.
- Chapman, J. W., & Tunmer, W. E. (1995). Development of young children's reading self concepts: An examination of emerging subcomponents and their relationship with reading achievement. *Journal of Educational Psychology*, 87(1), 154-167.
- Cheung, G. W. (2008). Testing Equivalence in the Structure, Means, and Variances of Higher-Order Constructs With Structural Equation Modeling. *Organizational Research Methods*, 11(3), 593-613.
- Chiou, H.-J. (1995). *The estimation of reliability, validity, and method effects with invariance in the multipopulation-multitrait-multimethod design using hierarchical confirmatory factor analysis*. Unpublished Doctoral Dissertation, University of Southern California.
- Chiou, H.-J., & Hocevar, D. (1995). *Examination of population-invariant construct validity in the Multipopulation-Multitrait-Multimethod design*. Paper presented at the 1995 Annual Convention of the American Psychological Association, New York, NY.
- Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research*, 18, 115-126.
- Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, 31, 33-42.
- Clogg, C. C. & Haritou, A. (1997). The regression method of causal inference and a dilemma confronting this method. In V. R. McKim & S. P. Turner (Eds.), *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences* (pp. 83-112). Notre Dame, IN: University of Notre Dame Press.
- Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics*, 13, 261-281.

- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd edition). Mahwah, NJ: Erlbaum.
- Comrey, A. L. (1970). *Manual for the Comrey Personality Scales*. San Diego: Educational and Industrial Testing Service.
- Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, 34, 35-46.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in the psychological literature. *Psychological Bulletin*, 52, 281-302.
- Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, 105, 317-327.
- Diamantopoulos, A. (1994). Modeling with LISREL: A guide for the uninitiated. *Journal of Marketing Management* (Special Issue on Quantitative Techniques in Marketing), 10: 105-136.
- Diamantopoulos, A., & Siguaw J. A. (2000). *Introducing LISREL: A Guide for the uninitiated*. Thousand Oaks, CA: Sage.
- Fisicaro, S. A. (1988). A reexamination of the relation between halo error and accuracy. *Journal of Applied Psychology*, 73, 239-244.
- Fletcher, T. D. (2005) The effects of parcels and latent variable scores on the detection of interactions in structural equation modeling. (Doctoral dissertation, Old Dominion University) *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 66(5-B), 2872. 14
- Fletcher, T. D. & Perry, K. M. (2007, April). A comparison of parceling strategies in structural equation modeling. Paper presented at the 22nd Annual Conference of the Society for Industrial and Organizational Psychology, New York, NY.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobserved variables and measurement error. *Journal of Marketing Research*, 18, 39-50.
- Fornell, C., & Yi, Y. (1992). Assumptions of the two-step approach to latent variable modeling. *Sociological Methods & Research*, 20 (1), 291-320.
- Freedman, D. A. (1987). As others see us: A case study in path analysis. *Journal of Educational Statistics*, 12, 101-128.
- Freedman, D. A. (1997). From association to causation via regression. In V. R. Mckim & S. P. Turner (Eds.), *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences* (pp. 113-161, 177-182). Notre Dame, IN: University of Notre Dame Press.
- Gould, S. J. (1981). *The Mismeasure of Man*. New York: Norton.
- Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (1994). Role of parental motivational practices in children's academic intrinsic motivation and achievement. *Journal of Educational Psychology*, 86(1), 104-113.
- Guilford, J. P. (1962). Creativity: Its measurement and development. In J. J. Parnes & H. F. Harding (Eds.), *A source book for creative thinking*. New York: Scribners.
- Hagtvet, K. A., & Nasser, F. M. (2004). How well do item parcels represent conceptually defined latent constructs? A two-facet approach. *Structural Equation Modeling*, 11, 168-193.
- Hair, J.F. Jr., Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R.L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Prentice-Hall.

- Hall, R. J., Snell, A. F., & Singer-Foust, M. (1999). Item parceling strategies in SEM: Investigation the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, 2, 233-256.
- Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances*. Baltimore, MD: John Hopkins University Press.
- Hays, W. L. (1994). *Statistics*. Fort Worth, TX: Holt, Rinehart & Winston.
- Hershberger, S. (1994). The specification of equivalent models before the collection of data. In A. von Eye and C.C. Clogg (Eds.), *Latent variables analysis* (pp.68-108). Thousand Oaks, CA: Sage.
- Hocevar, D. & El-Zahhar, N. (1992). Cross-cultural differences in test anxiety: Establishing transliteral equivalence. In K. A. Hagtvet (Ed.), *Advances in test anxiety research*, (Vol. 7, pp. 48-61), Star.
- Hoelter, J. W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods and Research*, 11, 325-344.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945-970.
- Hoyle, R. H. (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage.
- Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 158-176). Thousand Oaks, CA: Sage.
- Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112, 351-362.
- Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 76-99). Thousand Oaks, CA: Sage.
- Hu, L., & Bentler, P. M. (1999). *Cutoff criteria for fit indexes in covariance structural Equation Modeling*, 6(1), 1-55.
- Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Education Psychology*, 16, 131-152.
- Ilgen, D. R., Barnes-Farrell, J. L., & McKellin, D. B. (1993). Performance appraisal process research in the 1980s: What has it contributed to appraisals in use? *Organizational Behavior and Human Decision Processes*, 54, 321-368.
- James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.
- Jöreskog, K. G., & Goldberger, A. S. (1972). Factor analysis by generalized least squares. *Psychometrika*, 37, 243-260.
- Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social science*. (pp. 85-112). New York: Academic.
- Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70, 631-639.
- Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Lang (Eds.), *Testing structural equation models* (pp. 294-316). Newbury Park, CA: Sage.

- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8.14: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.
- Jöreskog, K. G. & Sörbom, D. (1996). *LISREL 8: Users' reference guide*. Chicago: Scientific Software International.
- Jöreskog & Sörbom (1996a). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.
- Jöreskog & Sörbom (1996b). *PRELIS 2: User's reference guide*. Chicago: Scientific Software International.
- Kaiser, H. F., Hunka, S., & Bianchini, J. (1971). Relating factors between studies based on different individuals. *Multivariate Behavioral Research*, 6, 409.
- Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.
- Kaplan, D., Harik, P., & Hotchkiss, L. (2000). Cross-sectional Estimation of Dynamic Structural Equation Models in Disequilibrium. In *Structural Equation Modeling Present and Future: A Festschrift in honor of Karl G. Jöreskog*. (pp. 315-339). Lincolnwood, IL: Scientific Software.
- Kenny, D. A., & Kashy, D. A. (1992). The analysis of the multitrait-multimethod matrix by Confirmatory factor analysis. *Psychological Bulletin*, 112, 165-172.
- Kim, S., & Hagtvet, K. A. (2003) The impact of misspecified item parceling on representing latent variables in covariance structure modeling: a simulation study. *Structural Equation Modeling*, 10, 101-127.
- Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain respective parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement*, 54, 757-765.
- Kline, R. B. (1996). Eight-month predictive validity and covariance structure of the Alcohol Expectancy Questionnaire for Adolescents (AEQ-A) for junior high school students. *Journal of Studies on Alcohol*, 57, 369-405.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford Press.
- Koopmans, T. C., Rubin, H., & Leipnik, R. B. (1950). Measuring the equation systems of dynamic economics. In T. C. Koopmans (Ed.), *Statistical inference in dynamic economic models* (pp. 53-237). New York: Wiley.
- Landis, R.S., Beal, D. J., & Tesluk, P.E. (2000). A comparison of approaches to forming composite measures in structural equation models. *Organizational Research Methods*, 3, 186-207.
- Lee, S., & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research*, 25, 313-334.
- Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods*, 4, 192-211.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151-173.
- Lomnicki, Z. A. (1967). On the distribution of products of random variables. *Journal of the Royal Statistical Society, Series B*, 29, 513-524.
- Long, J. S. (1983). *Confirmatory factor analysis*. CA: Sage.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA:

- Addison-Wesley.
- Lowe, N. K., & Ryan-Wenger, N. M. (1992). Beyond Campbell and Fiske: Assessment of convergent and discriminant validity. *Research in Nursing and Health*, 15, 67-75.
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in application of covariance structure analysis. *Psychological Bulletin*, 114, 185-199.
- MacCallum, R. C., Roznowski, M., Mar, M., & Reith, J. V. (1994). Alternative strategies for cross-validation of covariance structure models. *Multivariate Behavioral Research*, 29, 1-32.
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, 51, 201-226.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, 30, 41-62.
- MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding, and suppression effect. *Prevention Science*, 1, 173-181.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheet, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7, 83-104.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99-128.
- MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, 39, 384-389.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.
- Marcoulides, G. A. & Schumacker, R. E. (1996). *Advance structural equation modeling: Issues and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin*, 97, 562-582.
- Marsh, H. W. (1988). Multitrait-multimethod analyses. In J. P. Keeses (Ed.), *Educational research methodology, measurement and evaluation: An international handbook*. Oxford, Pergamon.
- Marsh, H. W., & Hocevar, D. (1988). A new, more powerful method of multitrait-multimethod analysis. *Journal of Applied Psychology*, 73, 107-117.
- Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, 13, 335-361.
- Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement*, 15, 47-70.
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantive meaningful distinction or artifacts? *Journal of Personality and Social Psychology*, 70, 810-819.
- Marsh, H. W., Hau, K., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33, 181-220.
- Maruyama, G. M. (1997). *Basics of structural equation modeling*. Thousand Oaks, CA: SAGE Publications, Inc.

- McDonald, R. P., & Marsh, H. M. (1990). Choosing a multivariate model: Noncentrality and goodness-of-fit. *Psychological Bulletin*, 107, 247-255.
- McDonald, R. P., Parker, P. M., & Ishizuka, T. (1993). A scale-invariant treatment for recursive models. *Psychometrika*, 58, 431-443.
- McDonald, R. P. (1997). Haldane's lungs: A case study in path analysis. *Multivariate Behavioral Research*, 32, 1-38.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- McDonald, R. P., Ho, M. R. (2002). Principles and practice in reporting structural equation analysis. *Psychological Methods*, 7, 64-82.
- Meade, A. W., & Kroustalis, C. M. (2006). Problems with Item Parceling for Confirmatory Factor Analytic Tests of Measurement Invariance. *Organizational Research Methods*, 9, 369-403.
- Meehl, P. E. (1959). Some rumination on the validation of clinical procedures. *Canadian Journal of Psychology*, 13, 102-128.
- Meehl, P. E., & Waller, N. G. (2002). *The path analysis controversy: A new statistical approach to strong appraisal of verisimilitude*, 7(3), 283-300.
- Meinert, C. L. (1986). Monographs in epidemiology and biostatistics (Vol. 8): *Clinical trials: Design, conduct and analysis*. New York, NY: Oxford University Press.
- Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, 29, 177-185.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). Washington, DC: American Council on Education and National Council on Measurement in Education.
- Mitchell, T. R., (1985). An evaluation of the validity of correlational research conducted in organizations. *Academy of Management Review*, 10, 192-205.
- Mosier, C. I. (1951). Batteries and profiles. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 764-808). Washington, DC: American Council on Education.
- Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.
- Mulaik, S. A. (1986). Toward a synthesis of deterministic and probabilistic formulations of causal relations by the functional relation concept. *Philosophy of science*, 52, 410-430.
- Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105, 430-445.
- Mulaik, S. A. (1995). The metaphoric origins of objectivity, subjectivity and consciousness in the direct perception of reality, *Philosophy of Science*, 62, 283-303.
- Mulaik, S. A. & Millsap, R. E. (2000). Doing the four-step right. *Structural Equation Modeling*, 7, 36-73.
- Murphy, K. R., Jako, R. A., & Anhalt, R. L. (1993). Nature and consequences of halo error: A critical analysis. *Journal of Applied Psychology*, 78, 218-225.
- Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52, 431-462.
- Nasser, F., & Wisenbaker, J. (2003). A Monte Carlo study investigating the impact of item parceling on measures of fit in confirmatory factor analysis. *Educational and Psychological Measurement*, 63, 729-757.
- Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 92-105). Washington, DC: American Psychological Association.

- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill, Inc.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill, Inc.
- Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, England: Cambridge University Press.
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (4th. ed). New York: Thomson Learning, Inc.
- Please, N. W. (1973). Comparison of factor loadings in different populations. *British Journal of Mathematical and Statistical Psychology*, 26, 67-89.
- Podsakoff, P., & Organ, D. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, 12, 531-544.
- Raine-Eudy, R. (2000). Using structural equation modeling to test for differential reliability and validity: An empirical demonstration. *Structural Equation Modeling* 7(1): 124-141.
- Raykov, T., & Penev, S. (1999). On structural equation model equivalence. *Multivariate Behavioral Research*, 34, 199-244.
- Raykov, T., Marcoulides, G. A. (2000). *A first course in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Raykov, T., Marcoulides, G. A. (2001). Can there be infinitely many models equivalent to a given covariance structure model? *Structural Equation Modeling*, 8, 142-149.
- Raykov, T. (2002). Analytic estimation of standard error and confidence interval for scale reliability. *Multivariate Behavioral Research*, 37, 89-103.
- Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy*, 35, 299-331.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114, 552-566.
- Reynolds, C. R., & Harding, R. E. (1983). Outcome in two large sample studies of factorial similarity under six methods of comparison. *Educational and Psychological Measurement*, 43, 723-728.
- Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research*, 30(3), 359-383.
- Rogers, W. R., & Schmitt, N. (2004) Parameter recovery and model fit using multidimensional composites: a comparison of four empirical parceling algorithms. *Multivariate Behavioral Research*, 39, 379-412.
- Rosenberg, M. (1986). *The logic of survey analysis*. New York: Basic Books.
- Sampson, C. B., & Breunig, H. L. (1971). Some Statistical Aspects of Pharmaceutical Content Uniformity. *Journal of Quality Technology*, 3, 170-178.
- Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In *American Statistical Association 1988 proceedings of the business and economics section* (pp. 308-313). Alexandria, VA: American Statistical Association.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors on covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 399-419). Thousand Oaks, CA: Sage.

- Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Erlbaum.
- Sharma, S., Durand, R. M., & Gur-Arie, O. (1981). Identification and Analysis of Moderator Variables, *Journal of Marketing Research*, 18, 291-300.
- Smith, H. F. (1957). Interpretation of adjusted treatment means and regression in analysis of covariance. *Biometrics*, 13, 282-308.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 290-312.
- Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. *Sociological Methodology*, 16, 159-86.
- Sobel, M. E. (1987). Direct and indirect effects in linear structural equation models. *Sociological Methods and Research*, 16, 155-176.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229-239.
- Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Springer, M. D., & Thompson, W. E. (1966). The distribution of products of independent random variables. *SIAM Journal on Applied Mathematics*, 14, 511-526.
- Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square tests. *Psychometrika*, 50, 253-264.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.
- Stelzl, I. (1986). Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research*, 21, 309-331.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics*. Needham Heights, MA: Allyn and Bacon.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th Ed.). Boston, MA: Allyn & Bacon.
- Tanaka, J. S., & Huba G. J. (1989). A General coefficient of determination for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, 42, 233-239.
- Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Newbury Park, CA: Sage.
- Tatsuoka, M. M. (1988). *Multivariate analysis: Techniques for educational and psychological research*. New York: Macmillan.
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago, IL: University of Chicago Press.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement equivalence literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-69.
- Vandenberg, R. J., & Scarpello, V. (1991). Multitrait multimethod validation of the satisfaction with my supervisor scale. *Educational and Psychological Measurement*, 52, 203-212.

- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 56-75). Thousand Oaks, CA: Sage.
- Williams, L. J., & Holahan, P. J. (1994). Parsimony-based fit indices for multiple-indicator models: Do they work? *Structural Equation Modeling*, 1, 161-189.
- Williams, F. (1980). *Creativity Assessment Packet: Manual*. East Aurora, NY: DOK Publishers.
- Wolfe, L. M. (2003). The introduction of path analysis to the social sciences, and some emergent themes: An annotated bibliography. *Structural Equation Modeling*, 10, 1-34.
- Wright, S. (1918). On the nature of size factors. *Genetics*, 3, 367-374.
- Wright, S. (1921). Correlation and causation. *Journal of Agriculture Research*, 20, 557-585.
- Wright, S. (1960). Path coefficients and path regressions: Alternative or complementary concepts? *Biometrics*, 16, 189-202.
- Yuan, K.-H., Bentler, P. M., & Kano, Y. (1997). On averaging variables in a CFA model. *Behaviormetrika*, 24, 71-83.

英汉术语对照表

adjusted GFI AGFI	baseline model 基准模型
completed mediation effects 完全中介效应	correlated measurement error 相关误差
absolute fit 绝对拟合	biased estimator 偏估计值
completely standardized solution 完全标准化估计值	Correlation 相关
admissible solution 可接受解	calibration sample 测定样本
composite reliability; CR; r 组合信度	Covariance 协方差
Aggregate 聚合	categorical variable 类别变量
conceptual model 概念模型	covariance structure analysis 共变结构分析
Akaike information criterion; AIC	causal modeling 因果模型
Akaike 讯息指数	covariance structure modeling 共变结构模式
concomitant variable 共存变量	Causality 因果性
alternative index 替代指数	Critical N; CN 关键样本指数
confidence interval; CI 置信区间	Causality 因果结论
alternative model 替代模型	Cronbach's α 内部一致性信度系数
confirmatory factor analysis; CFA 验证性因素分析	central chi-square distribution 中央卡方分布
alternative model competition 替代模式竞争比较	cross-causal model 跨因果模式
congeneric parcels 同源法	chi-square difference test 卡方差异检验
alternative models approach 替代模型取向	cross-validation 复核效化
Consistent Akaike information criterion; CAIC	classical test theory 古典测验理论
Akaike 一致讯息指数	data reduction 资料缩减
Analysis of Variance; ANOVA 方差分析	coefficient of correlation 相关系数
Constant 常数	deductive approach 演绎取向
asymptotic consistency 渐进一致性	coefficient of reliability 信度系数
constrained parameter 限定参数	degree of parsimony 简效程度
Asymptotic Distribution Free; ADF 渐进分布自由法	Collinearity 共线性
construct validation 构念有效化	dependent variable; DV 因变量
asymptotic unbiasedness 渐进不偏性	common variance 共同变异
control variable 控制变量	direct effect 直接效应
auto-correlation 自我相关	Communality 共同性
Convergence 收敛	directional relationship 直接关系
Average Variance Extracted; AVE; r_v 平均变异萃取量	comparative-fit index CFI 指数
correct weight matrix 校正权数矩阵	

- discriminant validity 区分效度
- competing models 竞争模型
- distorter variable 逆转变量
- distribution free 分布自由
- Disturbance 干扰项
- effect decomposition 效应分解
- Eigenvalue 特征值
- Elliptical distribution theory; EDT 椭圆分布理论法
- endogenous variables 内生变量
- Equivalent 等值
- equivalent model 等值模型
- exogenous variables 外源变量
- expected cross-validation index; ECVI 期望交叉效度指数
- exploratory factor analysis; EFA 探索性因素分析
- factor analysis 因素分析
- factor loading 因素载荷
- factorial invariance 因素恒等性
- factorial validity 因素效度
- feedback loop 回溯关系
- final solution 终解
- finite casual lag 有限因果性延宕
- first-order CFA 初阶验证性因素分析
- first-order factors 初阶因素
- fit function 拟合函数
- fitted covariance matrix 拟合共变量矩阵
- fitted residuals 拟合残差
- fixed parameter 固定参数
- free parameter 自由参数
- full latent model 完全潜在模型
- full matrix 完全矩阵
- full-information techniques 完全讯息技术
- general linear model 一般线性模型
- general structural equation model 一般化 SEM 模型
- generalized least squares; GLS 一般化最小平方方法
- goodness-of-fit index; GFI 拟合度指数
- hierarchical CFA; HCFA 阶层验证性因素分析
- hierarchical measurement model 阶层测量模式
- Hierarchy 阶层性
- higher-order factor 高阶因素
- hybrid modeling 统合模型
- hypothetical model 假设模型
- improper solution 不适当解
- incremental fit 增量拟合
- incremental fit index; IFI 增量拟合指数
- independent cluster basis 独立丛集策略
- independent model 独立模型
- independent variable; IV 自变量
- indicators 指标
- indirect effect 间接效应
- inductive approach 归纳取向
- information matrix 讯息矩阵
- instrumental variable 工具变量
- instrumental variable method; IV 工具性变量方法
- inverted information matrix 倒置讯息矩阵
- isolated uniqueness procedure 误差分离法
- item parceling 题项组合
- item-to-construct balance approach 项目构念平衡法
- iterative estimation 迭代估计
- just-identified 充分识别
- lag model 延宕模式
- latent variable 潜在变量
- limited-information techniques 有限讯息技术
- listwise deletion 全列排除法
- loose replication strategy 宽松复核
- manifest variables 外显变量
- Mardia multivariate kurtosis coefficient Mardia 多变量峰度系数
- matrices of factor loadings 因素载荷系数矩阵
- maximum likelihood; ML 最大概似法
- mean structure 平均数结构
- measured variable 测量变量

- measurement invariance 测量恒等性
 measurement model 测量模式
 mediated variable 中介变量
 method effect 方法效应
 MI index 修饰指数
 Minimum Fit Function Chi-Square
 最小拟合函数卡方值
 model generating approach 模型产生取向
 model generation 模型的产生
 model identification 模型识别
 model modification 模式修饰
 model selection 模型选择性
 model specification 模式界定
 model stability 模型稳定性
 model-fit evaluation 模型拟合评估
 model-implied or predicated correlation 模型衍生相关
 moderate replication strategy 温和复核
 multi-indicators multi-causes model MIMIC 模型
 multidimensional measurement 多维测量
 multidimensionality 多维性
 multi-group invariance 多样本恒等性
 multi-group structural equation modeling
 多样本结构方程模式
 multitrait-multimethod matrix; MTMM
 多重特质多重方法矩阵法
 Multivariate 多变量
 multivariate analysis of variance 多变量方差分析
 multivariate statistics 多变量统计
 naming fallacy 命名陷阱
 nested model 嵌套模型
 non-admissible 不可接受解
 noncentrality parameter, 非中央性参数
 non-centrality parameter; NCP 非集中性参数
 non-directional relationship 非直接关系
 non-normed fit index; NNFI 非正规拟合指数
 non-positive definition 非正定
 nonrecursive model 非递归模型
 non-saturated model 非饱和模式
 normed chi-square 正规卡方值
 normed fit index; NFI 正规拟合指数
 Null B Rule 虚无 B 矩阵法则
 observed correlation 观察相关
 observed variables 观察变量
 one-step modeling 一阶段模式
 operational definition 操作型定义
 operational model 操作性模型
 orthogonality rule 直交法则
 overall test 整体检验
 over-identified 过度识别
 paired deletion 配对排除法
 pallet indicators 平行指标
 parameter 参数
 parameter comparison 竞争比较
 parameter estimation 参数估计
 parcel scores 组合分数
 Parsimony Goodness-of-fit Index PGFI 指数
 partial cross-validation 部分复核效化
 partial latent model 部分潜在模型
 partial mediation effects 部分中介效应
 path analysis 路径分析
 path coefficient 路径系数
 Post-hoc test 事后检验
 precedence rule 前置原则
 principle of parsimonious 简效原则
 proper solution 适当解
 pure indicator 纯化指数
 qualitative variable 质性变量
 quantitative variable 量化变量
 radial parceling 辐射组合法
 random assignment 随机分布法
 random error 随机误差
 recursive model 递归模型
 recursive rule 递归法则

regression analysis 回归分析

Reification 抽象概念实质化

Reliability 信度

reparameterization method 再参数化策略

reproduced matrix S 导出矩阵

rescaling method 再量尺法

restricted measurement model 限制性测量模式

Robustness 强韧性

root mean square error of approximation; RMSEA
平均概似平方误根系数

root mean square residual; RMR 残差均方根指数

sample realitication 样本的依赖性

saturated model 饱和模式

second-order factor 二阶因素

simultaneous equation modeling 联立方程模型

smoothing procedure 平滑程序

solution completely standardized; SC 完全标准化解

specification searches 界定搜寻程序

spurious correlation 拟似相关

spurious relationship 虚假关系

squared multiple correlation, SMC 多元相关平方

standard error 估计标准误

standard error of sampling 抽样标准误

standard solution; SS 标准化解

standardized residuals 标准化残差

standardized root mean square residual; SRMR
标准化残差均方根指数

standardized solution 标准化估计值

starting values 起始值

Statistic 统计量

statistical model 统计模型

structural coefficient 结构系数

Structural Equation Modeling, SEM 结构方程模式

structural regression models 结构回归模型

structure model 完整结构模式

structure model 结构模式

suppressor variable 压抑变量

systematic error 系统误差

test for partial invariance 部分恒等性检验

the numbers of data points; DP 测量数据数

the ratio of improvement in noncentrality
非中央性改善比

theoretical model 理论模型

third-order factor 三阶因素

tight replication strategy 严格复核

total effect 整体效应

tracing rule 轨迹法则

triangular inequality condition 三角不均等条件

t-Rule t 法则

two-stage least squares; TSLS 二阶段最小平方法

two-stage procedure 二阶段估计程序

two-variable effects 双变量效应

unidimensional measurement 单维测量

uni-dimensionality 单维性

unique variance 独特变异

Uniqueness 测量残差或独特量

Univariate 单变量

unrestricted measurement model 未限制测量模式

un-standardized residuals 非标准化残差

unweighted Least Squares; ULS 未加权最小平方法

validation sample 效度样本

validity extension 效度延展性

validity generalization 效度类化

Variable 变量

vector of disturbance 干扰向量

vector of endogenous latent variables 潜在内生变量向量

vector of exogenous latent variables 潜在外源变量向量

weighted least-squares; WLS 加权最小平方法

χ^2/df 卡方自由度比